

Where next on arms control?

After the US election six weeks from now, the United States may resemble the Soviet Union in wishing to reduce arms spending for economic reasons. Where should it put its energies?

WHILE Mr Mikhail Gorbachev continues to dazzle with fresh proposals for arms control — limited demilitarization of the Pacific is the latest, last week — his counterparts in the West are inhibited from responding. The next president of the United States (whoever he may be) is wrapped up with the election on 8 November, most other Western politicians are similarly mesmerized and President Ronald Reagan shows little sign of planning to end his term with an historic flourish. So much is predictable, even understandable. In an election in which one of the few issues to have been raised is that of which of the principal candidates is most resolute in matters of defence, open advocacy of arms control could easily be misrepresented. It is to be hoped that the timing of the election, and the time it will take the new president of the United States to get to grips with his job, does not mean that an unprecedented opportunity is lost.

That the present offers remarkable opportunities is not in dispute. Several independent influences, of which Mr Gorbachev himself is one, are in conjunction. In contrast with his predecessors over seventy years, he is openly persuaded that the balance must be adjusted between military spending and domestic consumption in the Soviet Union. It is not merely that it would be beneficial if the military establishments (which apparently function effectively) could be redeployed in the interests of civilian industry (which does not), but that the purpose of military defences whose cost can only impoverish those whom they are intended to protect is, to say the least, dubious. Obviously, like any candidate in the US election, Mr Gorbachev could not make this view over-explicit without serious political risks, domestically and elsewhere. What we have, instead, is a torrent of proposals for arms control which are usually imaginative, sometimes mutually contradictory and which, taken together, would be beyond the capacity of the world's diplomats to negotiate. The result is that, while now more delicately placed domestically than for fifty years (witness Armenia and the Baltic States), the Soviet Union has given potential adversaries the conviction that it is genuinely seeking a military accommodation.

Balance

Another influence is the changing balance of the world outside. The development of Japan in a quarter of a century is a shining demonstration to people elsewhere that there is no necessary correlation between economic and military power: traditionally, the former has begotten the latter, but the historical correlation has tempted many to believe that military power can make nations prosperous (which has not been true since the Middle Ages, when armies were indistinguishable from pirates). Much of the enthusiasm in Europe for the promised open market of 1992 stems from emulation of Japan, while US envy (even fear) of Japan's economic success will provide the new president with a further incentive to balance his government's budget, no doubt at the expense of military programmes, the largest single discretionary item in the US budget. Soon, it seems, Mr Gorbachev will have quite a band of followers seeking to spend less on arms.

To cheer them up, there is the treaty on intermediate-range missiles (INF) signed last year (and now ratified) to show that

tangible steps forward are indeed possible. To the extent that the treaty will remove the threat of attacks on European targets by nuclear missiles owned and operated by the United States and the Soviet Union, it is plainly worth having. But, like other treaties, the value of INF lies as much in what it promises. The fact that the United States and Soviet Union have accepted, as part of the process of verifying compliance, arrangements for on-site inspection going much further than would have been contemplated just a few years ago, has lifted the spirits of those looking for other kinds of treaties — a ban on chemical weapons production, for example. In a different context, the existence of the treaty has reinforced West German ambitions for an agreement on short-range nuclear weapons. But INF is not an unmixed blessing: if nothing substantial follows, the its benefits will be eroded by the determination of European states to compensate for the removal of intermediate-range weapons by deploying other kinds, air-based cruise missiles for example.

Economy

So what can the next step be? By January next, it will surprisingly be apparent that the last six of the eight Reagan years will have left an unprecedented and rich if complicated legacy of uncompleted arms control agreements. The negotiations at Geneva on strategic missiles may be described either as half-finished or half-undone: the two sides have settled on a limit of 6,000 warheads and other details, but questions such as the use of sea-based cruise missiles and the interpretation of the anti-ballistic missile treaty remain. Work has also begun, but barely, on a chemical weapons treaty at Geneva. The Mutually Balanced Force Reduction talks at Vienna have been abandoned after fifteen years, but will be replaced in November by more tangibly defined negotiations on conventional weapons systems. (If the driving force for arms control has become economy, this is where there is money to be saved; when even rifle bullets have their own microprocessors, the cost of defence will be ruinous.) A new president may be readily persuaded, next January, that there is plenty of material with which to respond to what Mr Gorbachev has been saying.

Sadly, while each of these putative agreements (and there are many others) would be worthwhile in itself, none of them will in isolation avoid the problems occasioned by INF, in particular the temptation of signatories to restore their abandoned strength by other means. Nations required to live with a quota of nuclear warheads (6,000 is in any case more than enough) will put extra effort into the design of more effective delivery vehicles. Limiting numbers of battle tanks in Europe will boost work on other means of capturing targets. In any case, the treaties likely to emerge from these protracted negotiations are certain (like INF) to be cumbersome legal documents which, of necessity, will be open to later dispute.

That is why the best response to Mr Gorbachev's demand for action would be a simplifying response, one designed as much to simplify the process of arms control as to augment security in the short run. That is why there is a case for returning to almost forgotten issues. At the Moscow summit earlier this year, it was agreed that the verification procedures of the threshold test-ban



treaty (never ratified) would be re-negotiated so that the treaty could be put successfully to the US Senate; why not, at the same time, reduce the limit on explosive power from 150,000 to 10,000 tonnes of TNT equivalent and, at the same time, persuade France and China to join up? Or why not dust off some of the proposals for creating demilitarized zones in Central Europe, of which little has been heard since the early 1960s? □

Record-breakers vanish

This year's Olympics will probably see few records broken. Why not improve the supply?

THE quadriennial trauma of the world's sensibilities represented by the Olympic Games now unfolding in Seoul, South Korea, should not be allowed to conceal the difficulties with which this peripatetic institution will sooner or later have to grapple. For the supply of record-breaking performances is steadily decreasing, so that the organizers must be alarmed that there may not be in future years enough of them to persuade 200,000 visitors to faraway cities and to keep 1,000 times as many people glued to their television screens for two weeks.

Appearances to the contrary notwithstanding, athletic performance is a physiologically determined phenomenon which, as time goes on, will be increasingly governed by statistical considerations. As time goes by, world records will become increasingly rare. For those who enter competitions for jumping high, or long, or for throwing pieces of wood or lead as far as possible, are presumably members of the general population with a natural aptitude for doing what the competition requires and who have been trained to perform consistently. (Much the same is true of those who compete at running quickly over predetermined distances, except that foot-races involve tactical considerations, on-line calculations of how other runners are performing, which complicates the problem.) They cannot escape the underlying statistical character of their work.

The natural performance of individuals may be plausibly supposed to be randomly distributed within the general population. The distribution will have a long tail towards the high-performance end and will be markedly skewed. (Some people can really jump, most can hardly get off the ground.) Similarly, the performance of untrained people with athletic pretensions will be statistically distributed, from one performance to another, and will again have a long tail towards the high-performance end. (Most performances will be mediocre, some will be remarkable.) While trainers would no doubt claim that their function is to improve on nature, the much more probable benefit of their endeavours is to make their charges perform more consistently.

What those who have travelled to Seoul should now appreciate, is that there must be, now and in the years ahead, a growing scarcity of world records to be witnessed. Especially for well-established competitions such as jumping, the potential record breakers are those with exceptional ability among the general population, which has now been well worked over. It must be expected that, from time to time, individuals outstanding by contemporary standards will be discovered, but of necessity those must be increasingly rare events. And the effect of training, by reducing the standard deviation of an individual's performance-distribution, will be to rob these competitions of the surviving element of chance.

The International Olympic Committee seems well aware of this dismal prospect, having this year attempted to import collective sports such as baseball into the programme. The surprise is that it has not sought to keep the world records flowing more readily by more accurate measurement of competitors' performances. High-jumps, which tend now to be callibrated in centimetres, might be enlivened if the measurements were made accurate to the nearest millimetre or even micrometre. □

Genome monitoring

The organization to monitor the human genome project should be welcomed.

THE gibe that a substantial part of the 3,000 man-years of effort required to sequence the whole human genome has already been spent on discussions of the project is manifestly unfair, but is nonetheless significant. That people are still talking, not just getting on with the job, is a pointer not so much to the technical complexity of the project as to its social and political ramifications. It makes good sense that the difficulties that lie ahead should be anticipated.

That is the spirit in which the quaintly named organization christened HUGO should be welcomed (see page 286). So far as can be told, the objective is that HUGO should be an independent international body of selfappointed people whose interest is to encourage the genome project, to monitor progress and to anticipate problems. While HUGO hopes to find some funds to spend on modest enterprises, they are bound to be a tiny fraction of what will need to be spent on the production of nucleotide sequences of DNA from human and other genomes. The trick that HUGO hopes to play is that of influencing the policies of governments and of international and national institutions. Success will depend on being smarter than the governments that will eventually have to pay for the project.

Perhaps the place to start is with a definition of what the human genome project consists of. The simple notion that it will suffice to take the chromosomes from one person's cells and sequence the DNA from end to end is of course absurd, witness the gibe "Whose DNA?" The potential benefits of a successful project in this field are epitomized by the prospect that it will become easier to identify genes implicated in genetic diseases, but that is a small part of the tale there may be one day to tell. It is surely even more significant that the ideal product of a sequencing project should illuminate in ways not otherwise possible the crucial questions about the functioning of aggregates of genes, and about their evolution. For that purpose, the sequence of a single genome will not suffice. The variable parts of the human genome will have to be described and understood.

The good reason why people have not so far buckled down to the task of sequencing is that the scale on which these questions must be answered is not yet understood. The moral, for governments, is that the human genome project is not exclusively human and not a once-and-for all project. Their response should be much as it has been so far — supporting technical developments that may accelerate and even automate routine sequencing, assisting with the discussion of problems, scientific and technical, and brooding about the policy questions that arise. But this phase should soon be coming to an end. What the world needs is a number of small laboratories primarily skilled at sequencing that will also serve as foci for planning the way ahead. HUGO is probably well placed to help them coordinate their work, avoiding the danger that the project might become a kind of international competition, as to the Moon.

The policy questions are more difficult. For example, there is something in the view, shared by many, that resources spent on sequencing will be taken from other worthwhile projects in biology; the best answer is to demonstrate that the sequencing project itself will yield equal benefits. The ethical issues are more shadowy. Plainly, difficulties would arise, for example, if people carrying deleterious genes were able to tell their own status by referring to a sufficiently detailed genetic map. Physicians rightly ask that people carrying Huntington's disease should not be so informed without counselling. The more common worry, that there may be something about the constitution of the human genome that could be misused, is probably ill-founded but must nevertheless be countered intelligently if the project is to succeed. If HUGO can do this, it will be well worthwhile. □

SLAC calls time out on Stanford linear collider

- Particle collisions disappoint
- Earlier mode of operation planned

Berkeley

THE attempt by the Stanford Linear Accelerator Center (SLAC) to build an innovative electron-positron linear collider (SLC) has stalled. After a series of technical problems, the new machine has been temporarily shut down, and the facility will revert to its earlier mode of operation, with separate electron and positron storage rings. The goal now is to make changes that will allow the accelerator complex to switch between the two modes.

SLC is a technically daring addition to the linear accelerator, from which it takes beams of electrons and positrons alternately to send them around separate curving arms to a collision point. It was begun in 1983 as a competitor to LEP, a more conventional circular electron-positron collider at CERN, the European Laboratory for Particle Physics near Geneva, Switzerland, and is the brainchild of Burton Richter, who succeeded Wolfgang Panofsky as director of SLAC in 1984.

The new machine was finished on schedule last year, but by then CERN had already won the race to find the W and Z⁰ particles of electroweak unification. Still, US high energy physicists hoped that SLC would be the first machine to produce the Ws and Z⁰s in large quantities, allowing their properties to be studied in detail.

But SLC has not worked as planned. The beams have proved difficult to focus and control, and the collision rate has been low. Meanwhile, LEP is scheduled for completion next summer, and SLAC's 2-GeV and 14-GeV storage rings, which depend on SLAC's linear accelerator for the generation of particle beams, have been lying idle for two years, longer than was originally expected, while all energies were devoted exclusively to the development of the SLC.

The decision to shift some emphasis back to use of the storage rings was driven by the need of those projects for new data, said SLAC spokesman Michael Riordan, coupled with the fact that the SLC would need to be shut down for further modifications.

Last month, after several weeks during which the operation of SLC had been plagued by breakdowns, Richter removed Rae Steining, head of the accelerator division, from his post and stepped in himself with a taskforce to tackle the collider's problems (see *Nature* 334, 554; 1988).

Following the recommendations of the

taskforce, SLAC will make several modifications to SLC this autumn, including the addition of 400 tons of magnets to help deflect background radiation generated by the spreading of the particle beams as they travel through the collider's opposing arcs.

Later in the year, SLAC physicists hope to bring SLC back on line, in conjunction with the storage rings. In theory, the accelerator can be used to operate the SLC for periods of 2–3 hours, interrupted by half-hour pauses needed to fill the storage rings. But the transition between modes promises to be difficult and time-consuming at first.

While this shift in priorities may seem destined further to impede the progress of the SLC, Riordan insists that SLAC is not "giving up"; he hopes to be running the SLC and the storage rings in tandem by next February, and to characterize the Z⁰ particle by the end of next summer. Such an accomplishment would help win the favour of Congress for future colliders based on the SLC prototype.

With hindsight, SLAC officials say the SLC design cut too many corners. But despite the handicaps of ageing equipment and cost cutting, they believe that they have proved the principle of using linear accelerators as a source of particles for colliding machines. "We have made small micron-size beams and brought them into collision", said Riordan. "But it would certainly help our case to make a few thousand Z⁰s before LEP comes on line."

Marcia Barinaga

Joe Faust

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Down the tube: a view into the damping ring vault from the tunnel linking the ring to the linear accelerator. This photograph was taken during construction, before the tunnel components had been installed.

US embarks on radon testing

Washington

THE US government last week declared radon gas a national health hazard and recommended that every home should be tested for the presence of the radioactive material. The announcement triggered a sell-out of domestic radon test-kits, a stock-market boom for companies that make them and a flood of telephone calls to the Environmental Protection Agency (EPA).

The latest surge of interest in the hazards associated with radon came after a survey commissioned by EPA of 11,000 homes in seven states which has shown that high radon concentrations are much more common than expected. Radon gas forms radioactive decay products that, when inhaled, can damage lung tissue. EPA and the Surgeon-General's Office now estimate that radon is the largest cause of lung cancer after cigarette smoking, and may cause 20,000 lung cancer deaths a year in the United States.

The dangers of radon have long been recognized in other than domestic settings, in uranium mines for example. Interest in the domestic hazards of radon was heightened by an influential reassessment of radon as a contributor to the natural radiation dose published earlier this decade (see *Nature* 290, 99; 1981) and, more recently, by the recognition of high radon concentrations in houses built on uranium-bearing rocks in the Appalachian region. These data suggest that modern buildings minimizing draughts are prone to high concentrations of the gas.

Risks of lung cancer from radon have been calculated from the incidence in people working in uranium mines, where radon levels can be very high. A detailed analysis was released in 1984 by the United Nations Scientific Committee on the Effects of Atomic Radiation. The US government did not begin to worry about radon in homes until 1986, however, when EPA published booklets explaining the problem and began national monitoring.

As the results came in from the EPA survey it was soon apparent that the problem is even more serious than had been thought. Levels of radon vary enormously — even between two houses in the same street. And high levels of radon are not necessarily confined to areas with particular kinds of rocks beneath them.

Given the scale of the problem, the joint conclusion of EPA and Surgeon General's Office is that people should test their own homes. Radon monitors are available at low cost. And high radon levels can usually be reduced quite simply and cheaply by increasing ventilation and sealing off the ground beneath a house.

Alun Anderson

Human genome organization is launched with a flourish

Washington

WHAT was five months ago merely an idea for an international council to promote collaboration on the mapping and sequencing of the human genome has now coalesced into a bona fide organization with money in its pocket.

HUGO, the Human Genome Organization, held its first council meeting on the 6 and 7 September in Montreux, Switzerland, and newly elected president Victor McKusick of Johns Hopkins University in Baltimore expects that HUGO will have opened offices around the world by the end of the year.

HUGO was born at a rump session of a Cold Spring Harbor meeting last April. According to McKusick, James Watson, director of the laboratory, Leroy Hood of the California Institute of Technology, Sydney Brenner of the Medical Research Council Laboratory of Molecular Biology in Cambridge and Kenichi Matsubara of Osaka University were HUGO's intellectual godfathers.

The European Molecular Biology Organization (EMBO) is the model HUGO intends to follow, and to that end it has incorporated in Switzerland. HUGO will be an extra-governmental organization, but will depend on government contributions for its existence. It will give fellowships, conduct mapping workshops and issue annual reports.

Another possible function is to serve as a clearing-house for information about the growing number of international groups focusing on the genome project.

The HUGO Council proposed in Montreux that the organization should also start planning for the day when international centres will be set up around the world to do the immense — but mostly routine — task of sequencing identified fragments of DNA. McKusick says it is premature to consider establishing such centres now, but that they will be needed and that it would help to establish HUGO's identity if it were to take the lead in planning for them.

HUGO has established five areas of special interest: data banks, physical mapping/sequencing, other species, ethics and human disease. McKusick says that more cooperation is particularly to be encouraged between those involved in constructing genetic maps of the genome and those involved in physical mapping and sequencing efforts. He also believes that HUGO will have to be conscientious in addressing the ethical issues raised by the mapping effort.

If nothing else, HUGO would command attention based on the stature of its membership. Five of the 42 individuals

on the HUGO Council are Nobel laureates (Watson, Jean Dausset, Renato Dulbecco, Walter Gilbert, Francois Jacob), and the rest represent a *Who's Who* of molecular genetics. The council membership is international: 12 are from the United States, 7 from Britain, 5 from Germany, 4 from France, 3 from Japan, 2 each from Canada, Holland and Sweden, and one each from Australia, Greece, Italy the Soviet Union and Switzerland.

The Soviet member, Andrei Mirzabekov from the Institute of Molecular Biology of the Soviet Academy of Sciences attended the Montreux meeting.

HUGO does not yet have a fixed

budget, but is aiming for several million dollars a year to support its activities. A fund-raising drive is currently underway, directed by Harvard University's Gilbert, who was elected treasurer of HUGO. His selection for the post has an ironic tinge, as Gilbert has recently abandoned attempts to raise capital to start the Genome Corporation, a private effort to complete the mapping and sequencing project (see *Nature* 332, 387; 1988).

Gilbert is cautious about HUGO's fund-raising, but others are confident that several hundred thousand dollars can be quickly raised to set up HUGO offices in North America, Europe and Japan.

Brenner is credited with having come up with the name for HUGO, but with McKusick as the organization's president, some call it "Victor's HUGO".

Joseph Palca

Forty-year project endangered by NERC's regrouping

London

RESTRUCTURING of the research councils in Britain continues with the Natural Environment Research Council (NERC) announcing last week its decision to concentrate oceanographic research at one main centre on the campus of the University of Southampton. The move will cost £35 million and the government has not yet agreed to provide the funds. If the plan goes ahead, Southampton will become "the capital of oceanography in Europe" says Dr John Woods, head of the council's



NERC's Deacon Laboratory and the University of Southampton will have the use of the International Maritime Institute's Brave Challenger.

marine sciences directorate. Facilities there will match those of the Marine Biology Laboratory at Woods Hole in Massachusetts and the Scripps Institute of Oceanography in California.

Projects to be run jointly by the council and the university last week received support worth £1 million from a private trust, the International Maritime Institute, founded recently with the aim of promoting links between oceanographic institutions worldwide. The institute will provide

its high-speed vessel *Brave Challenger* for experiments to measure ocean storms, plankton distribution and to study estuarine environments.

Elsewhere, council support for plankton monitoring is to be withdrawn. A project which has monitored plankton in the north-east Atlantic Ocean and in the North Sea for 40 years continuously is to be the latest victim of the financial problems of the council. It is run by the council's marine laboratory at Plymouth, where researchers are dismayed by the sudden news that the project may be ended within a month. Their data are used in the United States and Canada as well as in several European countries and the researchers are confident they could find alternative funding, given time. The council is expected to make a decision soon on when the project will end.

Christine McGourty

■ The secretary of the research council most severely hit by the government's budget cuts, the Agricultural and Food Research Council, last week criticized government policy on science. Speaking at a meeting of the Science Policy Research Group, Professor William Stewart urged the government to withdraw funds from near-market areas of research "more appropriate for direct industry support", in favour of raising support for the research councils from the present "absolute minimum". But Professor Stewart also criticized the complacent attitude of many in the publicly funded sector towards the training for the commercial exploitation of science, saying that "all science undergraduates should take a compulsory course on the commercial exploitation of science". □

Japanese fellowships go begging despite \$2,000-a-month pay

Tokyo

Do young US scientists not want to come to Japan? Or are they simply unaware that there are now one hundred postdoctoral fellowships available for visits to Japan's government-run research institutes?

Figures released last week reveal a severe shortage of applicants for new Japanese fellowship schemes, originally designed to counter US criticism of the imbalance in exchange of scientists between the two countries.

During the re-negotiations of the Science and Technology Agreement, completed only in June, the United States lambasted Japan because far more Japanese researchers go to the United States than US researchers to Japan. Japan's response was a series of fellowship schemes whose supply outstrip demand.

Thus Japan's Science and Technology Agency (STA) devised a new postdoctoral fellowship scheme offering 100 fellowships a year. Half of the places were reserved for US scientists, with the remainder divided among 10 countries (Australia, Canada, West Germany, Finland, France, Italy, Netherlands, New Zealand, Sweden, Switzerland and Britain) and the European Community.

The Japanese Prime Minister Takeshita also arranged for the Foreign Ministry to donate \$4.8 million to the US National Science Foundation (NSF) to set up a fellowship fund for about 50 to 100 US scientists to make long-term visits to Japan, while the Ministry of Education, Science and Culture announced 100 "new" fellowships, again half going to the United States.

The ministry's fellowships, which are administered by the Japan Society for Promotion of Science (JSPS), have been squeezed out of existing bilateral programmes with Britain, France and West Germany by cutting the duration of the European fellowships from two years to one. For years, the European and particularly the British fellowship schemes have suffered from a chronic lack of applicants.

The new US fellowships schemes seem to be heading for the same fate. By last week, more than a month after the first deadline for applications (already extended), there were only four firm applicants for the US quota of 50 STA fellowships, according to Toyokazu Fukui of the agency's international affairs division. And there are only eight applications for the 50 fellowships available for US scientists under the JSPS scheme, according to Alexander DeAngelis, head of NSF's Tokyo office.

The conditions of the awards are unlikely to be to blame. The STA fellow-

ship, for periods of six months to two years, provides a tax-free monthly allowance of about \$2,000 (270,000 yen), plus additional payments to help with housing, travel and relocation. The JSPS fellowships offer similar conditions. The STA fellowships are tenable at any of over 100 government laboratories.

What has gone wrong? DeAngelis says it is too early to judge. He says the NSF office in Washington, which handles the US applications, distributed information on the fellowships to about 3,700 heads of departments at universities across the United States. He says that NSF has had more than 1,500 inquiries by letter or telephone, that the number of telephone calls is increasing and that he is optimistic that applications will pick up by the spring.

Part of the problem may be insufficient publicity. For years the Royal Society of London was able to fill only about half of the places under its scheme with JSPS. But after taking new measures to publicize the scheme in 1986 (*Nature* 323, 571; 1988),



applications began to rise and the scheme is now oversubscribed by more than two to one, according to John Grote, Science Officer at the British Council in Tokyo.

Inadequate notice and poor timing may also be to blame, according to DeAngelis. He says pamphlets describing full details of the STA fellowships were printed by STA only a little over a month ago, after the initial deadline of 1 August. Information may thus have been distributed to US universities when many researchers were on vacation. But Fukui says details of the fellowships were first described "informally" to western government officials in March and "officially" at the end of June.

STA officials are now worried. The agency has just applied for extra funds to boost the number of fellowships to 130 in fiscal year 1989 and the Ministry of Education has made a similar request to increase the JSPS fellowships to 130. But if this year's quotas cannot be filled by the end of December, when next year's budget must be fixed, the Ministry of Finance will be unwilling to provide the increases.

David Swinbanks

One-week wait

THE National Aeronautics and Space Administration (NASA) announced last Friday that the space shuttle *Discovery* would be launched on Thursday, 29 September, at about 10.00 am in Florida. The announcement was delayed until after Hurricane Gilbert had moved west and no longer threatened the Johnson Space Flight Center in Houston, Texas. D.L.

Ozone flyer

WEST Germany is to build the world's largest ultra-lightweight aeroplane for research into the chemistry of stratospheric ozone. Named "Egret", the plane will be constructed from carbon and glass fibre materials by a company that normally manufactures sailplanes. It will enter service next year and will be capable of flying to altitudes of 15 kilometres. Despite its size, its wing span will be relatively small — just 30 metres — to minimize drag. Crew will be a pilot and a scientist, who will be accompanied by 1,000 kilos of scientific equipment. J.N.

Two satellites up

JAPAN'S National Space Development Agency successfully launched a communications satellite last Friday aboard an H-1 rocket. When the Sakura 3b satellite is placed into geostationary orbit — in about 3 months — the way should be opened to commercial satellite services in Japan. A few days after the Japanese launch, Israel became the eighth nation capable of launching an indigenous satellite, orbiting a device principally designed for reconnaissance. Prime Minister Shamir called the success "a giant step making Israel a partner in the upper echelons in the technological world". □

Academy postpones

THE general meeting of the Soviet Academy of Sciences scheduled for the end of this month (27 September) has unaccountably been postponed until the end of October, according to members of the Academy. The general meeting, one of three held each year, was intended to reach a number of important decisions about the management of its own affairs in the light of the Party Conference at the end of June (see *Nature* 335, 101; 1988). □

Edinburgh Festival II

A NEW science festival will make its mark on the British public next year; to be held annually in Edinburgh, the festival will have much the same format as the British Association, though its director Howard Firth says it will be more community oriented. Funds of £85,000 have been raised from local authorities and commercial sponsorship; the target budget is £250,000. The theme of the first festival, to run from 3 to 12 April, is communications. C. McG.

Gorbachev rails at Siberian science and "asset strippers"

London

THE "scientific approach" to the development of the Soviet Union's remoter regions came in for a drubbing from Mr Mikhail Gorbachev during his visit last week to Krasnoyarsk, the Siberian city at the centre of a region rich in unexploited mineral and natural resources.

Speaking to an audience of scientists, economic planners, agricultural managers and workers, he condemned past policies which, in the name of creating "building sites of communism", had meant that "both nature is under threat and man's interests find themselves somewhere down the line at the lowest levels".

Among those responsible for this situation he singled out Academician Guri Marchuk, President of the Soviet Academy of Sciences, who, when head of the Siberian branch of the Academy had "failed to speak out and take a firm stance". A scientist's duty in such circumstances is, said to Mr Gorbachev, to "stand up straight and intervene at any state level, if what is necessary is not done".

The Krasnoyarsk *krai* (territory) is potentially one of the richest areas of the Soviet Union, possessing 50 per cent of Soviet coal reserves, 20 per cent of timber and hydroelectric resources, and a burgeoning non-ferrous metallurgical industry. No other region receives more capital investment, and the territory possesses academies, territorial production comp-

lexes and a planning department.

"One would think that a reasonable approach could be organized on that basis", Mr Gorbachev observed. Instead, plans for the development of the territory had gone badly wrong, production facilities had been built without housing or a service infrastructure, and hydroelectric plants supposed to provide the power base for development had caused considerable environmental damage.

Mr Gorbachev said that managers and administrators had behaved more like asset-stripping colonizers than developers concerned with the long-term potential of the country. Some hydro-plants had been built too cheaply, while the endemic Soviet failing of "gigantomania" had led the planners to support huge hydroelectric complexes, like those at Krasnoyarsk and Sayan-Shushenskoe, without asking whether smaller stations, sited in accordance with the natural courses of the rivers, would have served better.

Publicly chastising the failings of one region or sector as a timely warning to the country as a whole is a traditional Soviet method of launching a new policy drive. But Mr Gorbachev's speech to the Krasnoyarsk scientists seems particularly apposite. What seems to have moved Mr Gorbachev to his stern speech is, first and foremost, the mood of the workers, who are losing patience with the "scientific solutions" offered.

Vera Rich

Congress puts hot waste plans on ice

Washington

THE opening of the first permanent storage site for nuclear waste in the United States has been indefinitely delayed following a congressional hearing which, last week, revealed numerous doubts about the safety of the planned facility. The Waste Isolation Pilot Plant (WIPP) would have been one of the world's first purpose-built geological repositories.

Seven hundred million dollars have already been spent at the WIPP site in Carlsbad, New Mexico, where a two-thousand foot deep network of tunnels and chambers has been excavated in thick salt deposits. WIPP was due to begin receiving nuclear wastes from weapons production plants in October. Included for permanent storage would have been transuranic waste material contaminated with plutonium, whereas the Carlsbad plant was not intended to store high-level waste still generating heat.

WIPP has been plagued with problems ever since the New Mexico desert was selected for waste disposal in the early

1970s. Although the site had been selected for the stability and dryness of the salt deposits, chamber walls started to crack and creep and weep brine when the project was well underway, raising the prospect that barrels of waste might be crushed or corroded.

At last week's hearing the situation became even more complex when it emerged that some of the wastes are mixed with industrial solvents that could not be safely stored underground. Undocumented changes in design made it difficult to assess the overall safety of the facility and its capacity to resist earthquakes and fires. And legal disagreements between the federal government and the New Mexico government have left it unclear who has the final responsibility for licensing the plant.

The Department of Energy now says that WIPP will not begin receiving waste until scientists, Congress and the public are satisfied it is safe. The department cannot say how long that will take.

Alun Anderson

Student influx hits German universities

Munich

THE number of students crowding West German universities has once again increased, catching policy-makers by surprise. For years, politicians have been predicting a big fall in the number of new students. Instead, the latest figures, released last month by the West German Federal Statistics Office, show an 11.1 per cent increase at universities and other institutes of higher learning for the 1987-88 academic year.

The number of new students fell slightly in 1984 and 1985, encouraging finance ministers in the *Länder* (states) to sharpen the budget axe. (West German universities are funded primarily by the *Länder*.) But a few experts, such as the West German Rectors' Conference, warned that this was wishful thinking. The latest figures bear them out. The total number of students is close to 1,355,000, near the peak of 1983-84. The universities are now operating at about 150 per cent of capacity. The increasing length of courses of study in both humanities and natural sciences implies that the universities will be packed well into the 1990s (see *Nature* 333, 198; and 333, 792; 1988).

Education ministry officials in several *Länder* have outlined measures designed to relieve overburdened professors and facilities. In Bavaria, the proposed two-year university budget for 1989-90 includes expansion in several areas, but a spokesman admitted that the expansion targets set earlier in this decade foresaw a smaller number of students. Some universities, such as the University of Munich, which already has 62,000 students, simply cannot expand any further.

In Lower Saxony, spokesman Arpad Bogya said that special funds will be made available to universities to help them through this difficult period, but the amount — several million Deutschmarks — is too small to be of significant help.

The ministry officials explained the perplexing figures in terms of two phenomena: that university studies are becoming more popular among school leavers, and that people who have completed military service or practical training courses in preparation for entering careers have then opted for university studies as well. But most expect a sharp drop in the number of new students in two or three years at the latest. A spokeswoman in the Science Ministry in North Rhine-Westphalia explained that the number of 6-year olds beginning school there had dropped by half between 1970 and 1985.

Steven Dickman

US 'hands-off' forest policy challenged by Yellowstone fire

Boston

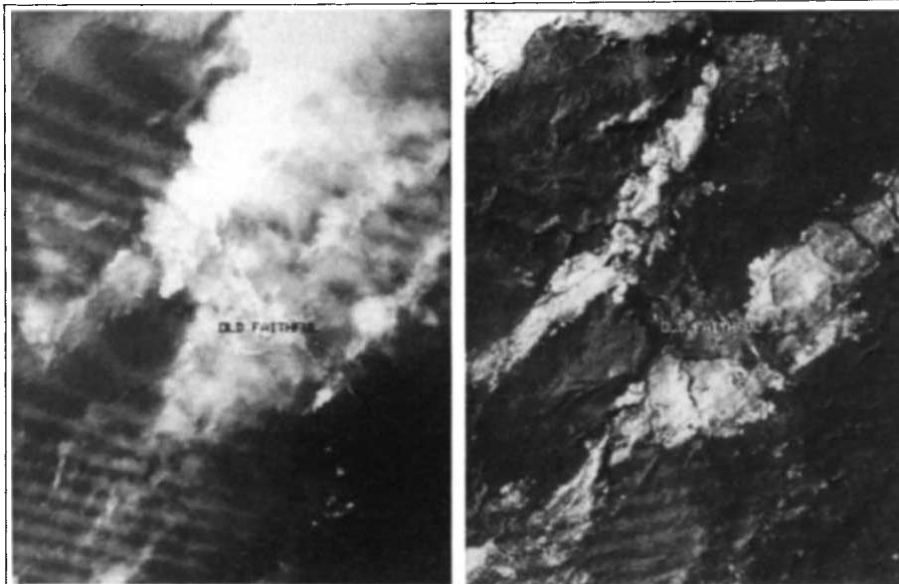
THE worst forest fires to strike the United States for many decades continued to rage last week at Yellowstone National Park in Wyoming and in neighbouring states, and have now ignited a vigorous debate about US policy for monitoring and managing forest fires in wilderness areas.

Including fires in Alaska, a total area of more than 4 million acres of forest has been burnt to the ground this year. At least 40 per cent of Yellowstone's 2.2 million acres has been razed. Since the early 1970s, the official US policy has been to let forest fires burn at national parks and

a major policy shift by the park service, is a readjustment of the balance on a difficult topic; congressional hearing, later in the year will debate it more fully.

At issue are fundamental questions about the nature and purpose of parklands in the United States. Most conservationists contend that the vast majority of these public areas are not frequently visited and should be kept in as natural a state as possible. Forest fires, they say, deplete the natural accumulation of burnable "fuel" made up of dry dead growth, and put fresh nutrients into the soil.

But others suggest that the heavy



Images of the fire crossing the Old Faithful community in the visible (left) and thermal infrared wavelengths. The burning area is seen as a light region in the thermal image. Photos taken on 9 September from a NASA ER-2 aircraft at 65,000 feet.

forests unless they clearly threaten human life or private property. Ecologists and conservationists have generally supported this "let-burn" policy, often called the "natural fire policy", on the grounds that fires are part of the natural process for rejuvenating forests. For a century before 1970, the United States had attempted to suppress fires in wilderness areas, which may have ultimately caused more destruction by allowing dry fallen wood to accumulate in dangerous amounts on the forest floors.

Under political pressure from the Wyoming delegation, Secretary of the Interior Donald P. Hodel announced last week that the National Park Service will now deviate slightly from its hands-off fire policy by planting new forests and vegetation in some park areas destroyed by this year's fires and by feeding some wildlife, such as elk and bison, to stop them migrating onto nearby private property.

Hodel's statement, while not indicating

recreational demands on key portions of the parks create a need for tighter management and control of the land. This group adds that the finite plots designated as wilderness areas should not be seen as self-regulating ecosystems, but need to be maintained to serve their function to society for recreation.

In practice, much of the debate may be inconclusive because of the semi-arid nature of many of the national forests in the United States, and the frequent occurrence of dry lightning storms. By almost all accounts, the accumulating woody fuels cannot be borne by the forest indefinitely. Once full-scale forest fires begin, it is unclear whether efforts at fire suppression are at all effective.

As present, some 30,000 fire fighters, including US Army and Marine troops, have been at work fighting the blazes. The effort has cost at least \$300 million so far and has claimed the lives of at least seven fire fighters. Recently the effort has been

Defence money for superconductors

Washington

IN a programme aimed specifically at fostering the technological development of high-temperature superconductors, the Defense Advanced Research Projects Agency (DARPA) is planning to distribute about \$30 million to nearly 40 industrial and academic research groups.

Although devices of some military consequence are the ultimate goal, DARPA's initial purpose is to improve the physical properties of the superconducting ceramics and to distinguish feasible from unrealistic applications.

Robert Shelton, a physicist at the University of California at Davis, praised the DARPA programme for its effort to coordinate the research talents of many different organizations, and for selecting a number of rather low-key but achievable technological goals, such as levitating bearings, magnetic sensors and superconducting coatings for radio-frequency cavities.

Shelton is associated with Ceracon Inc., of Sacramento, California, which has been given \$500,000 to adapt its high-pressure ceramic fabrication methods to the creation of dense superconductors with higher critical currents than the usual laboratory product.

Recipients of DARPA money met in Washington last week to compare notes and to give DARPA a chance to assess progress and suggest new directions and collaborations. According to Shelton, companies or universities with specific areas of expertise that fit in with DARPA's aims stand to profit from the programme, but those that turned up with vague ideas and some hope of a few hundred thousand dollars of federal money went away disappointed.

David Lindley

joined by a specially trained fire fighting contingent from Canada. Infrared scanners from NASA, flying in planes as high as 70,000 feet above the fires, have mapped a thermal picture of the fires to aid those trying to contain them.

Despite these efforts, most acknowledge that the campaigns have been largely futile. Some of the most serious property damage may have even been caused by fires started intentionally in an effort to create fire-breaks.

When the dry conditions and high winds favour a spreading forest fire, there apparently is not much that can be done. As Steven Whitney, director of the National Parks Program for the Wilderness Society, said, "you could have had wall-to-wall fire hydrants at Yellowstone and it probably wouldn't have made any difference".

Seth Shulman

Concern mounting over French government action on AIDS

Paris

THE four-month-old French government can no longer be in doubt that it will have to change the pace of national measures to combat the spread of AIDS. *La rentrée* — the return to work after the two-month holiday period — is always accompanied by an impatience to set everything to rights. But a series of independent broadsides in the press from eminent workers in the fields of public health and community care has pushed AIDS to the top of the pile on the health minister's desk.

The first public attack came from three eminent doctors. In the 9 September issue of the biweekly magazine, *Nouvel Observateur*, Leon Schwartzberg, whose career as Minister Delegate for Health in the new government lasted only nine days (see *Nature* 334, 461; 1988), Jean-Claude Chermann, who, with Luc Montagnier, first isolated the human immunodeficiency virus (HIV), and Paul Milliez, honorary dean of the Broussais-Hôtel-Dieu faculty of medicine in Paris, call upon those in power "who have not yet, it appears, appreciated the extent of the danger, to take action, at government level, to tackle screening for and information about AIDS".

"We ask the French public to accept generalized but voluntary screening. This alone will enable us to know who is carrying the virus and who is not", says their open letter. "It would be desirable that these tests be reimbursed by the Social Security. But, in order not to endanger the financial stability of the Social Security and to accelerate the process, we appeal to all French citizens who are able to, to pay for their own AIDS test at a cost of FF103 (\$16.50) per year".

The authors would like to see screening targeted "not only at populations whose behaviour puts them at risk, but also at pregnant women, those intending to marry, those undergoing surgery, indeed, all those who are sexually active and especially the young, who are particularly exposed".

The letter, in fact, calls for little that is not at present carried out, albeit haphazardly. Several maternity hospitals, on their own initiative, already advise expectant mothers to be screened. Screening for HIV is available to all on a voluntary basis in 400 laboratories and 113 screening centres, even if their infrastructure is often said to be wanting. The cost of the test is also refunded for those with Social Security insurance. But the Social Security system is running a huge deficit and would be unable to cope with massive use of state-reimbursed voluntary tests.

What is lacking is a high-profile co-

ordinated programme of information, counselling, care and targeted screening, as Daniel Defert, president of a leading national and voluntary action group, AIDES, points out in the daily newspaper, *Libération*. But Defert disagrees with Schwartzberg's suggestion that political motives alone are responsible for the absence of 'generalized' screening for HIV antibodies. "If screening has not been generalized, it is not because of a fear of impinging on individual freedoms, but for medical reasons, as researchers and clinicians repeatedly point out". Defert does, however, echo a growing feeling that successive governments have not grasped the nettle in their public health campaigns for fear of shocking public morals.

Defert agrees with Schwartzberg and his cosignatories that government action is

inadequate in the face of an annual doubling of AIDS sufferers in France.

Health minister, Claude Evin, has so far kept a low profile in the face of these attacks and is unlikely to announce government plans before October. By then he should have had time to digest the report, commissioned at the end of August and due in two weeks, to advise him what steps he should take (see *Nature* 335, 2; 1988).

Peter Coles

■ French government financing of AIDS research has also come under attack. Speaking at the Third International Conference on AIDS in Arusha, Tanzania, Luc Montagnier has again decried what he feels to be the low level of support for AIDS research. New grants for research on AIDS have dropped from FF100 million (\$16.13 million) last year to FF20 million for 1988. But in a Research Ministry reply, minister Hubert Curien explains that FF50 million of last year's allocation was not spent. This sum has been carried forward for 1988, bringing the total grant to FF70 million. □

Report cites unemployment as major health risk in Britain

London

FREE condoms, health warnings on alcoholic drinks, an absence of cigarette advertisements and mandatory rear seatbelts are just a few features of the new, healthier Britain that could be created if the government heeds the recommendations of an independent committee which has carried out a comprehensive study of the nation's health. In its report, published this week, existing health policies are severely criticized and increased funding for health promotion is called for.

The study was undertaken by an independent multidisciplinary committee sponsored by the Health Education Authority, the Scottish Health Education Group, the London School of Hygiene and Tropical Medicine and the King Edward's Hospital Fund for London.

Comparisons made with other European countries show that progress in public health in Britain has been neglected, inequalities within the country are glaring and the development of health promotion agencies has been uncoordinated, says the report. "There is an overriding need for the development of a coherent, long-term plan for the public health." That will include increased funding for national education programmes, the Sports Council and cancer screening services.

The tobacco, alcohol and confectionery manufacturers together spend almost £700 million annually on promoting products which represent hazards to health. In 1985-86, this represented more than 50 times the total budget of the bodies

responsible for health education.

Some specific recommendations are sure to arouse fierce opposition. The committee recommends that an annual levy be raised on the advertising of food, alcohol and tobacco, which would be used to promote good health. A 10 per cent levy on cigarettes would raise £10 million, it says. It also recommends new tar classifications — dangerous (under 5 mg per cigarette) and very dangerous (5-10 mg per cigarette).

It recommends warnings to consumers on what constitutes excessive or unsafe patterns of drinking; no further relaxation of the licensing laws and stiffer penalties for drinking and driving.

The report calls for a more coherent strategy to prevent AIDS and other sexually transmitted diseases. Condoms should be available on prescription from general practitioners, should be more widely available and should undergo mandatory quality testing.

One aspect of the report likely to embarrass the government is the link between health and socio-economic status, particularly with the unemployed. There are 1,500 extra deaths every year associated with unemployment for every million employed. The number of suicides among the unemployed is double that of the employed. This is the area in which major gains can be most readily achieved, the report says. "A reduction in unemployment is likely to have a major benefit on wide-ranging aspects of health."

Christine McGourty

Soviets put on impressive show

Moscow & Washington

LAST week's visit of US scientists and arms negotiators to the Soviet Semipalatinsk test site to witness an underground nuclear test was a success, according to US officials. Speaking in Moscow, C. Paul Robinson, chief US arms negotiator, said that talks would now go ahead in Geneva to try to ratify two nuclear test treaties before US President Ronald Reagan leaves office.

The two treaties are the Threshold Test Ban Treaty of 1974 and the Peaceful Nuclear Explosions Treaty of 1976. Both would limit underground tests to 150 kilotons but neither has been ratified, chiefly because of US fears that it would not be possible to detect violations.

The Semipalatinsk test follows by a few weeks an almost identical test at the US Nevada site (see *Nature* **334**, 641; 1988), where Soviet officials monitored the explosion of a US nuclear bomb. Of the two tests, the Soviets provided the closer view. Trailers full of monitoring equipment were set up two and a half miles from the 1,800-foot-deep borehole containing the nuclear bomb, in contrast to the US test where observers were kept thirty miles away. Before the Soviet bomb was triggered, the monitoring team was told to "bend their knees like skiers" according to Robinson. Even so, the force of the shock-wave, which blew the ground into an eight-foot-high bubble above the borehole, came as a jolt.

Data from the two tests are intended to resolve differences between the Soviet Union and United States over what kind of procedures will make it possible to monitor nuclear tests accurately. The US side has favoured the highly intrusive CORTEX technique that requires a cable to be placed in a hole alongside the bomb. The Soviets have argued that long-distance seismic monitoring is adequate.

First results from the Nevada test caused some embarrassment to the US administration when it was claimed that CORTEX showed the explosion to have violated the 150-kiloton unratified treaty limit. Traditional seismic methods apparently gave lower values closer to the expected yield. The results support the view of many scientists that all methods have some uncertainties and that seismic methods are adequate.

The United States and Soviet Union will now try to decide on the mix of CORTEX and seismic techniques that will satisfy both sides that the other is not cheating. The Soviet Union has been opposed to the use of CORTEX but now appears resigned to its inclusion.

Mary Manning & Alun Anderson

Harvard starts own marketing venture at "arms length"

Boston

HARVARD University plans to establish a limited partnership with a private investment consortium so as to be able to profit directly from the fruits of its own faculty's research.

The university will not be the first in the United States to undertake such an initiative, but its decision to play a role in the commercialization of faculty research marks a reversal of its well publicized policy of not taking a direct interest in commercial ventures springing from faculty research.

Almost exactly eight years ago, after much debate, Harvard officials rejected a similar plan to form a genetic engineering company with biochemistry professor Mark Ptashne. In a widely distributed report in 1980, Harvard president Derek C. Bok warned that such entrepreneurial ventures on the part of the university could compromise its objectives by "introducing into the very heart of the academic enterprise a new and powerful motive — the search for commercial utility and financial gain." Despite such warnings, the university's new plan seeks ultimately to identify research with lucrative commercial possibilities and establish small "spin-off" companies. The new arrangement, which will focus initially upon bio-

medical research at the Harvard Medical School and affiliated hospitals, involves an intricate scheme to keep the university at "arm's length" from financial decisions.

Specifically, the university plans to form a subsidiary, called Ion Inc., which will join a limited partnership of six to eight outside, institutional investors. This group will attempt to raise an initial \$30 million that will, in turn, be managed by another investment concern headed by Andre L. Lamotte, a former drug company executive recruited by Harvard Medical School for the job.

In exchange for providing start-up capital, the university will receive ten per cent of the profits from any new companies formed after the initial investment in them has been repaid. If the companies are sold, the university will be entitled to a tenth of the capital gains arising. The money raised by Harvard through the venture will be used to augment university support for basic research.

Officials at the university stress the goals of bringing products to market more rapidly and of "fostering a class of inventions that traditionally fall between the cracks". These are often too small and unformed to interest industry but too applied for government and foundation support.

Seth Shulman

Fetal tissue panel labours to beat a presidential ban

Washington

THE Human Fetal Tissue Transplantation Research Advisory Panel, meeting for two days last week at the US National Institutes of Health (NIH), failed to conclude its work on a report on the scientific, legal and ethical issues involved in such research. But the panel did endorse the basic premise that "the use of post-mortem fetal tissue for purposes of research and therapy that has been obtained from legally performed abortions is acceptable".

The panel was convened at the request of Health and Human Services Assistant Secretary Robert Windom, who ordered a moratorium on federal funds for fetal tissue research pending the outcome of the panel's deliberations. Arlin Adams, a retired federal judge, chaired the panel, with Kenneth Ryan of Harvard Medical School and LeRoy Walters of the Center for Bioethics at Georgetown University as cochairmen.

The most troublesome problem seems to be that of distinguishing between the use of fetal tissue and the manner in which

it is obtained. Inevitably, the panel was forced to go over some well-trodden ground on the ethical issues concerning abortion. But there was a consensus that new ground had been covered on the ethical questions raised by fetal tissue research, including what constitutes "informed consent" for its use — and who should have the authority to give it.

After two and a half days of deliberations, the panel decided that another meeting is needed to provide answers to the questions that Windom posed in a letter last March. The panel must conclude its work by the end of November, so that its views can be considered at a meeting of the NIH director's advisory committee in December.

Hanging over the session was the threat of a pre-emptive executive order by President Reagan banning all fetal tissue research (see *Nature* **335**, 197; 1988). The White House has said that the order is still being considered, but administration officials have made it clear that they do not feel bound to wait for the NIH panel to produce its report.

Joseph Palca

Orthodoxy and homoeopathy

SIR—The issue of whether basophils degranulate at infinite concentrations of IgE in a cyclic manner (*Nature* 333, 816; 1988) is of great scientific and clinical interest. Yet the hardly dispassionate whirl of experiments and 'counter-experiments' (all attempting to convince the reader of an even-handed perspective) strain the thesis that research can live comfortably in separation from its more secular environment of politics and personal opinion.

One might lament the apparent inability of experimentation to resolve the issue. The reader is left, if he has no prior opinion, much like an undecided voter at the polls — amid the confident exhortations of campaigners manning the doors. Perhaps a team of politicians and public-opinion pollsters could be sent to the laboratories involved to help out the undecided.

JAMES A. SCOTT

Massachusetts General Hospital,
Boston, Massachusetts 02114, USA

SIR—The recent hullabaloo about ignoring negative results in the homoeopathy study may be only the tip of the iceberg. Unfortunately, as many medical scientists are already sceptical about homoeopathy, they will find comfort in this — and will ask what else could be expected. They do so at their peril. Clearly, ignoring negative results may be a disease far more prevalent than has been suspected and may not exist only in laboratories that attempt to investigate homoeopathy. Negative results, important as they are, often find their way into the wastepaper basket. Could this not also indicate a problem with the editorial policies of scientific journals which very rarely publish such results? In view of the suspicion that fraud is generally on the increase in science, such audits (perhaps random and unexpected), might be a good idea in all laboratories, even those doing conventional scientific work. The results of such audits might surprise everyone.

MARK A. GILLMAN

South African Brain Research Institute,
Suite 9, Highlands House,
173 Louis Botha Avenue,
Orange Grove 2192,
Johannesburg, South Africa

SIR—Those of us who practise homoeopathy and are interested in its scientific investigation can take some small satisfaction at seeing the assertion, made by homoeopaths for many years, that extreme dilutions do have biological effects, enter the realm of scientific discourse. For too long this claim has been derided as beyond the bounds of possibility and unworthy even of consideration.

Thanks to you and the editors of certain other journals, the cat is now out of the bag. It will not be put back by insults or by a *a priori* refusal to believe that such a phenomenon is possible.

Nor yet will the controversy be resolved by your *post hoc* inquiry which was, by both accounts, hasty. Theatricals such as the taping of foil-wrapped randomization codes to the ceiling can hardly be described as established scientific procedure.

Your delegation included no biological scientists, so this was not peer review, while the very composition of your delegation implies that it had been decided, in advance, that the explanation of the reported results must be sleight-of-hand, or fraud, as you did not include anybody capable of distinguishing genuine results from a technical error made in good faith. One wonders how many biological scientists (or conjurers) would presume to deliver a verdict on five years' work in physics involving collaboration between six independent laboratories, after just five days' investigation at one laboratory.

Neither can extreme dilution effects be prohibited by scientific law. There has been talk of 'violations of fundamental scientific principles'. But the only specific victim proposed has been the law of mass action. If the phenomenon is physically rather than chemically mediated, as proposed by Benveniste's group, then this law is irrelevant. Certainly no fundamental principles, such as the laws of thermodynamics, are at stake.

The question of the action, or lack thereof, of extreme dilutions will not be resolved by polemics, but only by genuine application of the scientific method. Extended and accurate experiments must be carried out, with control, but without prejudice, followed by publication, criticism and repetition.

PETER FISHER

Royal London Homoeopathic Hospital,
Great Ormond Street,
London WC1N 3HR, UK

SIR—The appearance of a new journal always attracts interest and the publication on 30 June 1988 of the first issue of *New Approaches to Truly Unbelievable and Ridiculous Enigmas* was no exception. But I wonder how many readers were shaken (not to say stirred) by the subsequent publication of comment which indicates that editorial wisdom in that journal, though still detectable, has become infinitely diluted in an ocean of rhetoric? Perhaps science, like journalism, also has a silly season.

M.J. CLEMENS

Ham Cottage, Albourne Road,
Hurstpierpoint, West Sussex BN6 9ES, UK

From other letters

THE Benveniste findings deserve to be investigated in a calm, unpressured and unprejudiced environment, when they will probably vanish like N-rays and polywater. Your rushed and evidently prejudiced attempt to discredit them has probably killed any chance of such a deliberate assessment occurring.

ALEXANDER M. GRIMWADE

706 Cambridge Road,
Bala Cynwyd, Pennsylvania 19004, USA

BENVENISTE's report does provide some evidence that insufficient effort was made to eliminate systematic and subjective biases in the experiments. But is this sufficient to label the results a "delusion"? I think not. Neither, apparently, did the team, as they have attempted to further "build a case" against Benveniste by throwing in a number of *non sequiturs*. Why should they have been "dismayed" by who paid the researchers' salaries? So what if the experiment didn't work "all the time"? Show me one that does, and I'll show you one where somebody has been fudging.

SCOTT FINDLAY

University of Ottawa,
Biology,
30 Somerset E.,
Ottawa, Ontario, Canada K1N 6N5

WHY should one require any more or less caution, make more or less study design and execution demands on an experiment that does not fit into a current explanatory (and so theoretical) model than one that does? It could be argued that it is when results fit neatly into current scientific dogma that a study should be required to be more stringently executed than if they do not, lest in confidence we make lax assumptions and so overlook perhaps more important alternative explanations for the results. Lateral thinking is a skill scientists could benefit from a bit more of.

The importance of the article by Benveniste and coworkers is not in the results of the study itself but in the fact that it was published in *Nature*, rather than in the *British Homoeopathic Journal*. Homoeopathic and other journals have for years published the results of experiments demonstrating the activity of ultra-high dilutions. Many of these studies have been done in a scientifically rigorous fashion.

WAYNE B. JONAS

Office of the Surgeon General,
Health and Fitness Division,
5109 Leesburg Pike,
Falls Church, Virginia 22041-3258, USA

Letters submitted for Correspondence should be typed, double-spaced, on one side of the paper only. □

New particles from old data

A search for an explanation of positrons in heavy-ion collisions has led to firm evidence of light neutral particles and an explanation of a forgotten anomaly in the lifetime of the π^0 meson.

PALAEONTOLOGISTS are the people able most commonly to boast of having made important discoveries in material already worked over by others, but the identification of the Crab Nebula with a long-since exploded star seen by the Chinese shows that they do not have a monopoly of delayed discovery. But now, surprisingly, almost forgotten data have been turned to good use in particle physics, helping to confirm what appears to be the discovery of a new particle.

It is a curious tale. It is now five years since the first reports from the Heavy-Ion Research Laboratory (GSI) at Darmstadt that collisions between heavy ions were often followed by the emission of positrons with characteristic (but not particularly large) energy. The data have never been sharp enough to persuade everybody that the phenomenon is real, but the theoreticians have from the outset been eager to demonstrate that the positrons are real and interesting. One ingenious attempt to show that the observed positrons are nothing but particles released directly from the universal Dirac sea of electrons with negative energy appears not to hold water, but most excitement has centred on the possibility that the positrons observed are the decay products of light neutral particles, of which that known as the axion would be the most desirable.

While the theoreticians have been saying what the explanation is, some experimenters have been looking for confirmatory data. A Berkeley-Stanford collaboration produced some intriguing data towards the end of last year. But now M. El-Nadi and E. O. Bawady, from the Physics Department at the University of Cairo, have produced what seems to be telling evidence that there is indeed a phenomenon to be explained (*Phys. Rev. Lett.* **61**, 1271; 1988). Their demonstration is all the more appealing because they used techniques that first came into their own four decades ago.

Plainly it is not essential, in the study of the interaction between heavy nuclei, to build a huge detector full of electronics, install it near a heavy-ion accelerator and then patiently collect information about events of predefined character over a period of weeks or months. What El-Nadi and Bawady did was to expose a stack of nuclear emulsions (of the kind used for identifying heavy ions in the cosmic rays reaching the upper levels of the atmosphere) to a beam of ^{12}C or ^{22}Ne ions at the

Institute of Nuclear Research at Dubna, looking for the interaction patterns revealed by the tracks in the exposed emulsion after it has been developed.

What the authors have done is to scan their emulsions for cases of patterns in which a primary nucleus has been broken up into smaller pieces in a collision with unidentified nuclei in the emulsion, and then to look for cases in which at least one of the secondary particles interacts with a further nucleus in the emulsion to yield an electron pair — a positron and an ordinary electron.

The outcome is satisfyingly graphic. El-Nadi and Badawy show a representation of an interaction track in which a ^{12}C nucleus disintegrates into three α particles, one of which, in a further interaction, yields a proton, a pion and a kaon (both mesons, one strange) and, downstream of the second interaction by $46\text{ }\mu\text{m}$, the apex of a narrowly diverging pair of tracks apparently caused by particles with the characteristics of electrons. The inference is that some electrically neutral particle spans the $46\text{ }\mu\text{m}$ gap. Out of 2,600 primary interactions in the emulsion, the authors collected 13 examples of electron-positron pairs produced in this way.

As in all these circumstances, there is an intricate argument about the likelihood that some of the observed electron pairs are produced by other means than by a previously unrecognised particle. El-Nadi and Badawy note that electron pairs can be produced by energetic photons and that energetic photons result from the decay of electrically-neutral π^0 mesons. Luckily, the background expected from this source is less than a third of the events recovered, leaving El-Nadi and Badawy able to estimate that the mass of their neutral particle is within 1.15 electron-masses of 3.13 electron-masses and that its lifetime (at rest) is just over 10^{15} seconds.

Forgotten data enter the argument at two places, but some of wit is the authors' own; they explain that the emulsion track they show (described above) was first presented at a conference at Darmstadt in 1985. But the data presented from Cairo may explain why the first measurements of the lifetime of the π^0 meson were two orders of magnitude greater than the value accepted when accurate measurements became possible. That at least is the assertion of F.W.N. de Boer and R. van Dantzig from the University of Amsterdam, in an accompanying paper which

reads as if it is a referees' report turned into print (*Phys. Rev. Lett.* **61**, 1274; 1988), an inference supported by the submission dates. (The Amsterdam paper was first submitted after a revised version of the Cairo paper had been received.)

de Boer and van Dantzig do not quarrel with the Cairo group about the background to be expected from other processes for forming electron-pairs, but they plot measured lifetimes and masses of the neutral particles inferred for fourteen events, finding that they fall naturally into three clusters more widely separated from each other than their error bars.

This is where the forgotten data count, for the authors also note that the first measurements of the lifetime of the π^0 meson were made in essentially the method followed by El-Nadi and Badawy — C.F. Powell's group at the University of Bristol looked for pairs of electrons in nuclear emulsions exposed to cosmic rays, reckoning that these would all have been the products of photons themselves produced from π^0 mesons. By good luck, there are even detailed analyses of the kinematics of two of the tracks observed at Bristol, which fall ("amusingly", as the authors put it) right at the centres of two of the clusters in their diagram. The inference is that the first lifetime measurements were really measurements of a quite different neutral particle.

It is strange how convincing, no doubt irrationally, is this unexpected coincidence. If El-Nadi and Badawy are right, their neutral particles will account not just for the puzzle they set out to solve, but for another of which they appear to have been unaware. It is enough to drive people into the arms of Bayesian probability.

Indeed, there is more to say. de Boer and van Dantzig note that at least one of the masses corresponds closely with one of those singled out by the Stanford-Berkeley experiment last year. Cautiously, they agree that there is at least one, and possibly three, neutral particles which decay into electron-pairs, with masses of 1.1, 2.1 and 9.2 MeV (the first coincides with the Cairo estimate). But none of them appears to be an axion in the simplest sense. Perhaps pairs of axions with opposite polarity might be involved, but that is for the time being speculation. But nobody will now deny that the phenomenon needs further investigation. There will be a brisk demand for nuclear emulsion.

John Maddox

Protein-DNA interaction

No code for recognition

Brian W. Matthews

THIS has been something of a banner year for repressor-operator complexes. On page 321 of this issue¹, Sigler and colleagues describe the structure of *trp* repressor complexed with its synthetic operator. Other complexes involving the DNA-binding domains of λ -repressor² and of phage 434 repressor³ as well as 434 Cro protein⁴ have also now been determined. The structure at near-atomic resolution of another complex of 434 repressor was described last year⁵. It is an opportune time to take stock of the field and to consider implications for the future.

In their general features, the observed repressor-operator complexes vindicate the models for DNA-protein interaction⁶⁻¹⁰ that were developed following the crystal-structure determinations of Cro⁶, CAP¹¹ and λ -repressor⁸. It is perhaps useful to summarize the features that were anticipated from the earlier work (p.282 of ref. 12; see figure) and, at the same time, to flesh out the additional insights that have been provided by the latest results of Harrison, Pabo, Ptashne, Sigler and their co-workers.

Symmetry. The overall repressor-operator complexes have approximately 2-fold symmetry, with the 2-fold axis of the protein coinciding with the 2-fold axis of the DNA. Differences in sequence between the two halves of the operator site are now seen to be associated with subtle adjustments in the protein monomers².

DNA conformation. The DNA maintains a conformation that is essentially standard, right-handed, B-form. As was anticipated, some deformation of the DNA can occur during the formation of the complex. An interesting situation occurs for 434 Cro and 434 repressor headpiece in which case the same 14-base-pair operator adopts different conformations when bound by the respective proteins⁴.

Helix-turn-helix motif. As predicted¹³⁻¹⁵, each of the repressor proteins¹⁻⁵ contains the now well-known helix-turn-helix motif with the second α -helix located in the major groove of the DNA (see figure).

A contentious question that can now be resolved concerns the alignment of the helix-turn-helix motif on the DNA. Is the helix-turn-helix unit a 'fixed reading head' that is positioned on the DNA in the same way in different DNA-binding proteins? Model-building studies with λ -repressor, Cro and CAP had suggested that this would not be the case^{9,10,16}. On the other hand, 'helix-swap' experiments, in which parts of the recognition helices of different repressors were interchanged, suggested

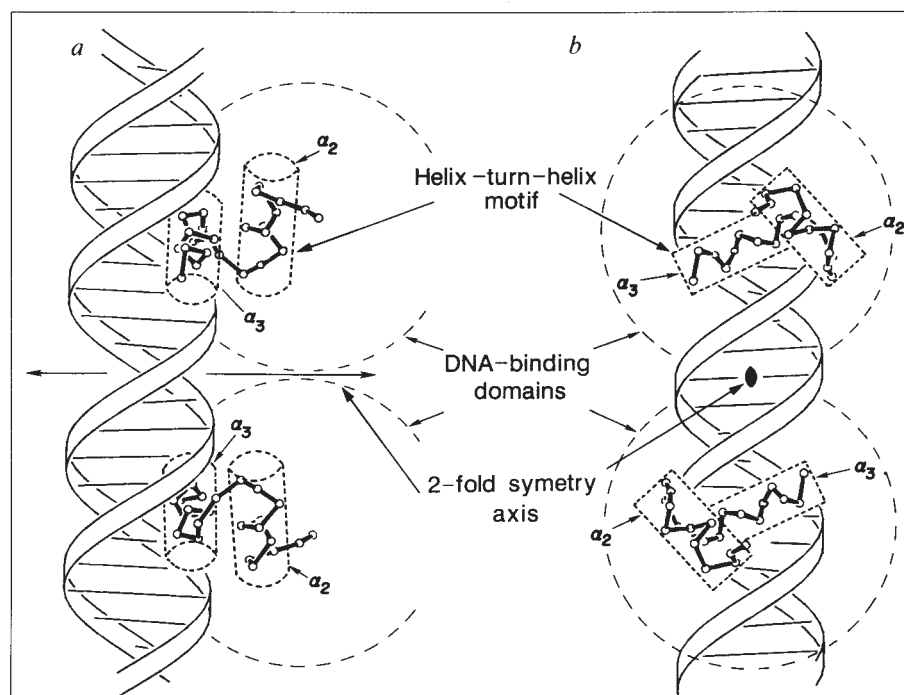
that the binding geometry of the recognition helix on the DNA is similar for many repressor proteins¹⁷. The complexes now available show that the geometry of binding is variable. Even in the case of 434 Cro⁴ and 434 repressor headpiece^{3,5}, two proteins with 50 per cent amino-acid sequence identity and very similar three-dimensional structures, and which are bound to the same operator, the respective helix-turn-helix units interact with the DNA in similar, but distinctly different, ways⁴.

Specificity. The early studies of Cro, CAP and λ -repressor suggested that these proteins recognize their specific binding sites on the DNA by multidentate hydrogen-bonding and van der Waals contacts between the side-chains of the protein and the edges of the base pairs exposed in the major groove of the DNA. With the exception of *trp* repressor described in this issue (see below), the complexes observed in the new work²⁻⁵ strongly support this mode of DNA-protein recognition. It is also now clear how the DNA structure adapts to enhance complementarity at the protein-DNA interface. Multiple interactions between

the protein and the phosphate backbone contribute to the overall affinity of binding; sequence-specific contacts with the exposed parts of the operator base pairs allow the repressor to discriminate between its operator site and other DNA sequences.

Direct or indirect readout? The complexes involving λ -repressor headpiece, 434 Cro and 434 repressor headpiece all display multiple contacts both to the DNA backbone and to the parts of the base pairs that are exposed within the grooves of the DNA²⁻⁵. Although deviations from uniform B-form are observed in the conformations of the DNA, the recognition of the specific operator sequences appears to occur primarily by direct readout. In other words, the DNA retains essentially B-type conformation and the protein directly reads the sequence information encoded within the grooves of the DNA^{7,9,10,18}.

Unexpectedly, however, the complex of *trp* repressor with its operator displays a different mode of interaction¹. In this case, there are no direct hydrogen bonds or non-polar contacts between the protein and those bases that have been shown *in vivo* to be important in recognition. In some cases, these bases are hydrogen-bonded to the protein via water molecules; in other cases they are exposed to solvent. Sigler and colleagues argue in this issue¹ that *trp* repressor provides an example of indirect readout. In other



The general mode of interaction presumed to occur in many DNA-regulatory proteins and now confirmed experimentally. The helix-turn-helix motif (shown here as helices α_2 and α_3) is common to all the repressor-operator complexes whose structures are known. In different complexes, the helix-turn-helix motifs bind in roughly similar, but distinctly different modes, showing that there is no simple code for DNA-protein recognition. *a*, A side view with the 2-fold axis of symmetry extending from left to right; *b*, view along the 2-fold axis. (From ref. 15.)

words, the operator sequence is recognized indirectly through induced changes in the geometry of the phosphate backbone that, in turn, permit the formation of a tight complex with the protein. It is known, for example, that changes in a central base pair of the operator of 434 repressor, which is not contacted by the protein, can alter the affinity of binding by sevenfold².

The *trp* repressor complex poses several questions. Is the distortion of the DNA backbone sufficient to allow the repressor to distinguish between specific and non-specific sites? This is difficult to assess at the present stage, in part because the observed deviations of the operator from uniform B-DNA appear to result from the cumulative effect of modest local distortions. It may also be significant that the preference of *trp* repressor for its operator site relative to random DNA sequences is about 10⁴-fold, which is somewhat less than many other repressor proteins. Because DNA is a highly charged molecule, the formation of any protein-DNA complex is likely to have a significant electrostatic component and therefore to be strongly salt-dependent. It is therefore desirable to confirm that a repressor-operator complex seen in crystals represents the sequence-specific complex that occurs in solution.

What is the role of water in the recognition of the *trp* repressor-operator? Sigler and colleagues argue¹ that water-mediated hydrogen-bonding could contribute to the direct readout of three base pairs within the operator sequence, but that such direct readout cannot explain the known importance of two additional base pairs that have been shown to contribute to recognition. The explicit need to consider bound water on the surfaces of both proteins and DNA adds another level of complexity to the recognition problem.

A universal recognition code? Is there a

code whereby certain DNA base pairs are recognized by certain amino acids? More specifically, are given operator sequences recognized by a simple set of rules in which certain amino acids in the helix-turn-helix motif confer specificity for certain bases? The answer, again, is no. First, it can now be seen directly that the helix-turn-helix motifs of different repressors do not align with their respective operators in identical modes. Second, the DNA-protein interface is seen to be very complex, with several side-chains sometimes contributing to the recognition of a single base². Finally, the conformation of a given DNA operator can change

ferently in response to the binding of different repressors.

It is very satisfying now to have in hand the structures of several repressor-operator complexes that vindicate the general principles of DNA-protein recognition that have been developed by many individuals during the past 20 years. But the full appreciation of the complexity and individuality of each complex will be discouraging to anyone hoping to find simple answers to the recognition problem. □

Brian W. Matthews is in the Institute of Molecular Biology, University of Oregon, Eugene, Oregon 97403-1229, USA.

Earth sciences

The origin of carbonatites

J. Gittins

CARBONATITES, rare magmatic carbonate-rich rocks, are usually associated with silica-undersaturated rocks and are found almost exclusively in continental cratons where they are often associated with rifting. Although first recognized early this century, carbonatites have had a controversial history because of the presumed impossibility of a carbonate-rich magma existing at petrologically reasonable temperatures and pressures. The paper by Wallace and Green on page 343 of this issue¹ offers an exciting insight into the possibility that carbonatite magma may be of direct mantle origin.

It was established in the 1960s that such magmas are possible, but there has been much debate since then as to whether carbonatite magma is developed directly by partial melting of mantle peridotite or by modification within the crust, and hence at much lower pressure, of a mantle-derived magma by fractional crystallization or liquid immiscibility. Much of the controversy has been summarized by LeBas² and by Twyman and myself³.

Carbonatites have unique chemical compositions characterized by extreme enrichment in rare-earth elements, in light relative to heavy rare earths, and in niobium, phosphorus, strontium and barium relative to their concentration in more common igneous rocks. Consequently, carbonatites are the sole commercially mineable source of niobium, the principal source of rare-earth elements, and a possible source of phosphorus. The last possibility is particularly relevant when there is strong pressure on the mining industry to reduce sulphur-dioxide emissions from smelting operations, as the sulphuric acid by-product can be disposed of by reacting it with the phosphorus mineral apatite to produce saleable super-phosphate fertilizer.

The mining of carbonatites allows the production of minerals that could not profitably be mined individually. An outstanding example is Phalaborwa in South Africa which produces copper (rare in carbonatites), vermiculite, magnetite (iron ore), apatite, uranium and zirconium. The isotope composition of carbonatites (strontium, neodymium, lead, carbon and oxygen) is particularly appropriate for investigating the Earth's mantle. Bell and co-workers^{4,5}, for example, have documented the history of the mantle under parts of Canada and East Africa using carbonatites.

Most information about the origins of the carbonatite magma has emerged almost incidentally from experiments on the origin of more common mafic and ultramafic rocks such as basalts. These 'phase equilibrium' experiments simulate the temperatures and pressures that exist at depths down to several hundred kilometres in the mantle and allow the experimenter to determine which minerals form stable assemblages and the composition of liquids in equilibrium with them. The discovery reported by Wallace and Green in this issue is an example: it was made while investigating the partial melting behaviour of a carbonate-bearing amphibole lherzolite in the pressure range 15–32 kbar, and suggest a means by which carbonatite magma could have been produced. An amphibolite peridotite (lherzolite) containing 0.3% H₂O and 0.5–2.5% CO₂, known to be present in the mantle, will produce through partial melting a carbonate liquid from which carbonatite can crystallize.

Most high-pressure studies of carbonatite magma genesis^{6,7} have produced liquids that are capable of crystallizing carbonate minerals but lack the sodium, potassium, silicon, aluminium, phosphorus and iron necessary for the range of

- Otwinowski, R.W. *et al.* *Nature* **335**, 321–329 (1988).
- Jordan, S.R. & Pabo, C.O. *Science* (in the press).
- Aggarwal, A.K., Rogers, D., Drott, M., Ptashne, M. & Harrison, S.C. *Science* (in the press).
- Wolberger, C., Dong, Y., Ptashne, M. & Harrison, S.C. *Nature* (in the press).
- Anderson, J.E., Ptashne, M. & Harrison, S.C. *Nature* **326**, 846–852 (1987).
- Anderson, W.F., Ohlendorf, D.H., Takeda, Y. & Matthews, B.W. *Nature* **290**, 754–758 (1981).
- Ohlendorf, D.H., Anderson, W.F., Fisher, R.G., Takeda, Y. & Matthews, B.W. *Nature* **298**, 718–723 (1982).
- Pabo, C.O. & Lewis, M. *Nature* **298**, 443–447 (1982).
- Lewis, M. *et al.* *Cold Spring Harbor Symp. quant. Biol.* **47**, 435–440 (1983).
- Weber, I.T. & Steitz, T.A. *Proc. natn. Acad. Sci. U.S.A.* **81**, 3973–3977 (1984).
- McKay, D.B. & Steitz, T.A. *Nature* **290**, 744–749 (1981).
- Ohlendorf, D.H. & Matthews, B.W. *A. Rev. biophys. Bioengng.* **12**, 259–284 (1983).
- Anderson, W.F., Takeda, Y., Ohlendorf, D.H. & Matthews, B.W. *J. molec. Biol.* **159**, 745–751 (1982).
- Sauer, R.T., Yocum, R.R., Doolittle, R.F., Lewis, M. & Pabo, C.O. *Nature* **298**, 447–451 (1982).
- Ohlendorf, D.H., Anderson, W.F. & Matthews, B.W. *J. molec. Evol.* **19**, 109–114 (1983).
- Ohlendorf, D.H., Anderson, W.F., Lewis, M., Pabo, C.O. & Matthews, B.W. *J. molec. Biol.* **169**, 757–769 (1983).
- Wharton, R.P. & Ptashne, M. *Trends biochem. Sci.* **11**, 71–73 (1986).
- Seeman, N.C., Rosenberg, J.M. & Rich, A. *Proc. natn. Acad. Sci. U.S.A.* **73**, 804–808 (1976).

silicates (micas, pyroxenes and amphiboles), phosphate and oxide minerals that are found in carbonatites together with the more common calcite, dolomite and ankerite. The liquid produced by Wallace and Green contains 3 % SiO_2 , 2 % Al_2O_3 , 5 % Na_2O , 0.4 % K_2O , 4 % (total) FeO and 3.5 % P_2O_5 , all of which allow for the crystallization of the non-carbonate minerals.

Despite the controversy over carbonatite magma having formed by fractionation of nephelinitic magma or by immiscible separation from nephelinitic or phonolitic magma, there is general agreement that all such magmas have experienced a protracted period of fractionation in the crust. The problem, then, has been to find a parental carbonatite magma that contains all the components necessary to have produced the range of non-carbonate minerals referred to above. The immiscibility hypothesis is becoming increasingly popular, but most studies to date have generated a fairly pure carbonate liquid^{8,9}.

The study by Wallace and Green offers a realistic parental magma from which the full diversity of carbonatite rock types may be developed by fractional crystallization, either of olivine, spinel and dolomite in the crust or mantle, or of the characteristic mica-pyroxene-amphibole-apatite-Fe-Ti oxide-carbonate suite within the crust. Extreme fractionation could have produced the highly alkalic carbonatite lavas currently erupting from the Tanzanian volcano Oldoinyo Lengai, whereas less extreme fractionation could have produced most calcitic and dolomitic carbonatites. But the high Cr_2O_3 content (3.2 %) in Wallace and Green's liquid further strengthens the case for the liquid being a pre-fractionation, primary, parental carbonatite magma, as carbonatites are not chromium-rich.

The next step is to determine whether the primary melt can also contain the necessary concentrations of rare earths and niobium to account completely for the range of characteristic carbonatite minerals. □

- Wallace, M.E. & Green, D.H. *Nature* **335**, 343-346 (1988).
- LeBas, M.J. in *Alkaline Igneous Rocks* (eds Upton, B.G.J. & Fitton, J.G.) 53-83 (Geol. Soc. Spec. Pub. **30**, 1987).
- Twyman, J.D. & Gittins, J. in *Alkaline Igneous Rocks* (eds Upton, B.G.J. & Fitton, J.G.) 85-94 (Geol. Soc. Spec. Pub. **30**, 1987).
- Bell, K. & Blenkinsop, J. *Geology* **15**, 99-102 (1987).
- Bell, K. & Blenkinsop, J. *Geochim. cosmochim. Acta* **51**, 291-298 (1987).
- Wyllie, P.J. in *Magmatic Processes: Physicochemical Principles* (ed Mysen, B.O.) 107-119 (Geochem. Soc. Spec. Pub. **1**, 1987).
- Eggler, D.H. *Austral. J. Earth Sci. Spec. Pub.* **14**, IV (1988).
- Freestone, I.C. & Hamilton, D.H. *Contr. Miner. Petrol.* **73**, 105-117 (1980).
- Kjarsgaard, B.A. & Hamilton, D.H. *Mineral. Mag.* **52**, 43-55 (1988).

J. Gittins is in the Department of Geology, University of Toronto, Toronto, Ontario, Canada M5S 1A1.

Pattern generators

Modulating a neuronal network

Eve Marder

How does the nervous system generate behaviours and how is performance modified to fit circumstances? To elucidate the neuronal mechanisms involved in producing a motor behaviour requires detailed analyses of the movements, of the neural circuits that produce them and of the cellular mechanisms generating the activity in the circuit. In practice, most experimental preparations fall short of this ideal, but Heinzel and Selverston¹⁻³ now achieve such a synthesis with their examination of the control of the lobster's stomach movements. From this and other recent work on similar invertebrate preparations, an understanding of the way neuromodulators work to select a particular output

pattern of nerve impulses is emerging.

There are three teeth in the lobster's stomach comprising the gastric mill, which grinds and shreds food. The movements of the teeth are generated and coordinated by a small group of neurons in the stomatogastric ganglion, a nerve centre situated in the dorsal artery on the surface of the stomach. In a technical *tour de force*, Heinzel^{1,2} used an endoscope to study the movements of the teeth in the living animal (Fig. 1), and he and Selverston³ have recorded intracellularly from up to seven neurons simultaneously to obtain a picture of the patterns of neuronal activity associated with the two modes in which the teeth operate: squeeze and cut-and-grind.

About half the 30 neurons in the stomatogastric ganglion are involved in controlling the rhythmic movements of the gastric mill (Fig. 2). The gastric rhythm has long fascinated aficionados of neuronal oscillators, because it was thought to be an emergent property of the circuit formed by these neurons, rather than being driven by an endogenous pacemaker neuron or neurons. We now know instead that rhythmic gastric-mill activity results from the plateau properties of many neurons in the circuit⁴, excitatory input from other ganglia^{5,6} and the synaptic connectivity among the gastric neurons themselves⁷.

Heinzel's endoscope investigations reveal that the gastric-mill rhythm results in movements of muscles and ossicles that produce complex three-dimensional displacements of the two lateral teeth and the single medial tooth. The animal can switch between the squeeze mode and the cut-and-grind mode on a cycle-by-cycle basis but produces only one or the other within

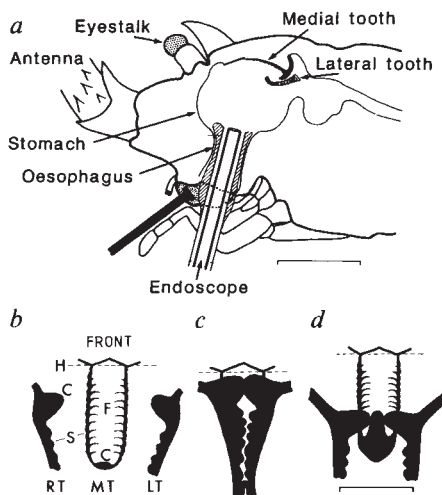


Fig. 1 a, Diagrammatic section of a lobster's head showing the stomach with the endoscope in position and two of the three teeth of the gastric mill. b - d, Schematic drawings of the three teeth in typical positions seen through the endoscope; b is the resting position.



100 years ago

NOTES ON METEORITES

AMONG the fragments in deep-sea deposits we distinguish granules . . . yellowish-brown, with a bronze-like lustre, and the surface, instead of being quite smooth, is grooved by thin lamellae. In size they never exceed a millimetre, generally they are about 0.5 millimetre in diameter; they are never perfect spheres; and sometimes a depression more or less marked is to be observed in the periphery. The lamellae which compose them are applied the one against the other, and have a radial eccentric disposition. It is the leafy radial structure (*radialblättrig*), like that of the *chondres* of bronzite, which predominates. The figure shows the characters and texture of one of these spheres magnified twenty-five diameters. From *Nature* **38**, 530; 27 September 1888.



a given cycle of neural activity. When the neuropeptide proctolin is injected into the circulation of an animal not producing gastric rhythms, rhythms similar to those recorded spontaneously are elicited. Low concentrations of peptide most often produce the squeeze mode, whereas high concentrations result in the cut-and-grind mode. As proctolin is present in neurons that form inputs to the stomatogastric ganglion⁸ and is a circulating hormone in crustaceans⁹, it will reach neurons in the stomatogastric ganglion under normal physiological conditions.

When proctolin is applied to preparations of the isolated stomatogastric ganglion *in vitro*³, the ganglion produces patterns of motor neuron activity that resemble closely those expected to underlie the effects of proctolin on the gastric-mill movements (Fig. 3); two of the neurons in the gastric network, the lateral gastric and the dorsal gastric, show strong plateau properties. This enhanced regenerative activity helps to account for the transformation in the motor patterns that occurs in the presence of proctolin, at least partially accounting for the changes in tooth coordination induced by the peptide.

The new work¹⁻³ exemplifies several themes emerging from studies of how small neural networks produce oscillatory outputs^{7,10-12}. First, the rhythmic output of the circuit derives both from the membrane properties of the component neurons and the synaptic interactions

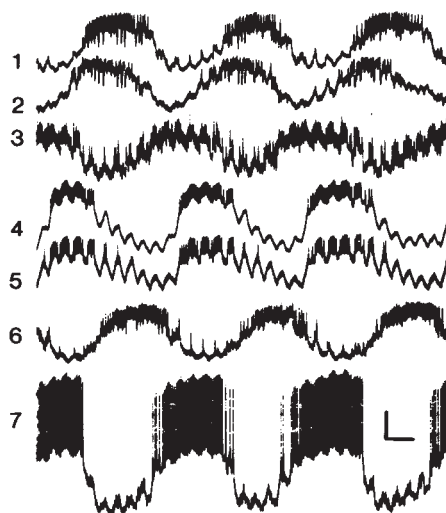


Fig. 2 Simultaneous intracellular recordings from neurons of the gastric system showing typical spontaneous gastric activity. Numbers, specific neurons, identifiable from preparation to preparation, some of which are discussed in the text. 1-6, motoneurons; 7, an interneuron. Scale bars: 1 s; 10 mV (from ref. 3).

among them^{3,7}. Second, many neurons have conditional oscillatory properties that contribute to the generation of network oscillations under some conditions but not others^{4,7,13}. Third, modulatory inputs control the expression of these conditional membrane properties, as well as change the strength of certain synaptic inputs^{11,12,14}.

Consequently, neural networks can provide variable outputs or can be 'tuned' or 'configured' in different ways¹⁰⁻¹². In the pyloric network of the stomatogastric ganglion, for example, each of the many different modulatory substances, including biogenic amines and several peptides, found in inputs to the ganglion¹⁵, produce a different form of the pyloric rhythm^{11,12,15}. In this case, several different substances can increase the burst frequency and amplitude of bursts in the anterior burster neuron, the pace-maker for the pyloric rhythm^{11,12}. But because these substances act differently on the other neurons of the pyloric network, the network response to each is different^{11,12,15}.

The phenomenon of 'reconfiguration' also occurs in *Tritonia*, where the sign of the synaptic interactions

between the dorsal swim interneurons is switched from inhibition in a non-swimming preparation to excitation during swimming¹⁶. In fact, comparing the features of many small neuronal networks¹⁶ shows that similar tasks can be performed by neuronal networks with different constructions¹⁶, and that a single neuronal network can produce many outputs^{11,12,15}.

The new understanding is thus that neuronal networks are not 'hard-wired' but are made of neurons whose membrane properties can be modified, connected by synapses whose strength, sign and time course can vary according to behavioural state and activity patterns in the network, to produce various output patterns. This makes recent advances in computational neuroscience especially exciting. Network models show that complicated tasks can be carried out by groups of simple 'neuronal-like elements' hooked together by 'synaptic rules'¹⁷.

The challenge is to elucidate how small neuronal networks work; how large networks of simple neuronal-like elements work; and how the complexity of neurons (conferred by their shape and ionic currents) and the richness of their connectivity define and constrain the possible functions of a neuronal network. So we need eventually to determine which properties of biological networks are not shared by networks of simple neuronal-like elements. Insights that have been gained from computer-modelled neuronal-like networks and studies of modulation in neuronal networks such as the stomatogastric ganglion should help to understand how large biological networks generate behaviours and process information. □

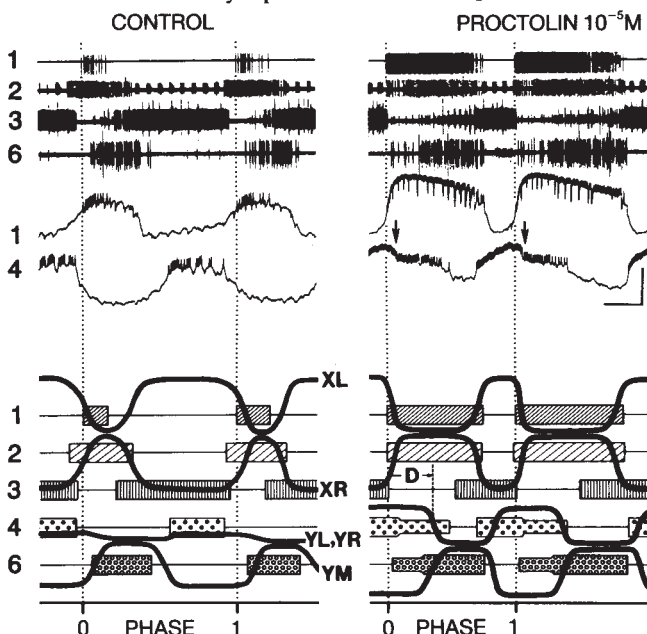


Fig. 3 The effects of proctolin on the gastric mill burst patterns. Top panels are extracellular and intracellular recordings from the gastric neurons showing the typical burst patterns in control (left) and proctolin-containing (right) saline. Numbers refer to the motoneurons shown in Fig. 2. Bottom panels are normalized block diagrams that show the time in the burst cycle when each neuron is active. The heavy black lines denote the movements of the teeth. The presence of proctolin alters the firing patterns of the individual neurons and so changes the coordination of the muscles that control the relative movements of the three teeth (from ref. 3).

1. Heinzel, H.-G. *J. Neurophysiol.* **59**, 528-550 (1988).
2. Heinzel, H.-G. *J. Neurophysiol.* **59**, 551-565 (1988).
3. Heinzel, H.-G. & Selverston, A.I. *J. Neurophysiol.* **59**, 566-585 (1988).
4. Russell, D.F. & Hartline, D.K. *J. Neurophysiol.* **52**, 54-73 (1984).
5. Robertson, R.M. & Moulins, M. *J. comp. Physiol.* **154**, 473-491 (1984).
6. Dickinson, P.S. *et al. J. exp. Biol.* **136**, 53-87 (1988).
7. Selverston, A.I. & Moulins, M. *The Crustacean Stomatogastric System* (Springer, New York, 1987).
8. Marder, E. *et al. J. comp. Neurol.* **243**, 454-467 (1986).
9. Beltz, B.S. & Kravitz, E.A. *J. exp. Biol.* **124**, 115-141 (1986).
10. Getting, P.A. in *Neural Control of Rhythmic Movements in Vertebrates* (eds Cohen, A.H., Rossignol, S. & Grillner, S.) 101-107 (Wiley, New York, 1988).
11. Marder, E. *et al. New Insights into Synaptic Function* (eds Edelman, G.M., Gall, W.E. & Cowan, M.W.) 305-327 (Wiley, New York, 1987).
12. Harris-Warrick, R.M. in *Neural Control of Rhythmic Movements in Vertebrates* (eds Cohen, A.H., Rossignol, S. & Grillner, S.) 285-331 (Wiley, New York, 1988).
13. Bal, T. *et al. J. comp. Physiol.* (in the press).
14. Nagy, F. *et al. J. Neurosci.* **8**, 2875-2886 (1988).
15. Marder, E. & Hooper, S.L. *Model Neural Networks and Behavior* (ed. Selverston, A.I.) 319-337 (Plenum, New York, 1985).
16. Getting, P.A. & Dekin, M.S. *Model Neural Networks and Behavior* (ed. Selverston, A.I.) 3-20 (Plenum, New York, 1985).
17. Hopfield, J.J. & Tank, D.W. *Science* **233**, 625-633 (1986).

Eve Marder is in the Biology Department, Brandeis University, Waltham, Massachusetts 02254, USA.

Radioactivity

Seeking non-exponential decay

P. T. Greenland

ONE of the most famous laws of physics is the exponential-decay law of radioactivity¹: the decay of any radioactive isotope is characterized solely by its time constant τ or equivalently by its half-life $t_{1/2} = \tau \log_2 2$. During a time t , $N_0 \exp(-t/\tau)$ nuclei will decay, where N_0 is the number of radioactive atoms at the beginning of the interval. This law has now been checked by Norman *et al.*² for extremely short periods, down to 0.01 per cent of a half-life, and long periods, up to 45 half-lives. In neither case was any deviation from exponential decay found. Why should so much effort be expended testing a law with such a venerable history? Because it is incompatible with the equally venerable theory of quantum mechanics³.

Classically, the exponential law holds because each radioactive atom in an ensemble decays independently and with a probability that is independent of time. As decay proceeds, there are fewer remaining radioactive atoms and the sample's activity falls. Exponential decay is seen in atomic and molecular decay processes as well as in radioactivity.

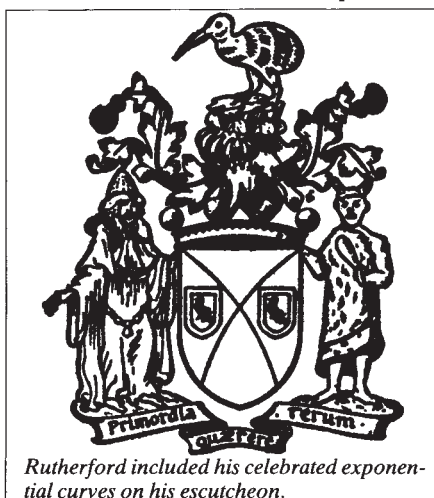
In quantum mechanics, the decay law of a prepared system should be derived from first principles. The system is described by a state vector Ψ which includes information on how many atoms have already decayed. A 'hamiltonian' operator $\hat{H}$ controls the dynamics of the system, which evolves according to the Schrödinger equation as $\Psi(t) = \exp(-i\hat{H}t)\Psi_i$, where Ψ_i is the state at time $t = 0$.

The decay law is obtained by evaluating the survival probability, that is, the component of the initial state remaining in wavefunction $\Psi(t)$ at time t (denoted $P(t) = |A(t)|^2 = |\langle \Psi_i | \Psi(t) \rangle|^2$). Certain physical restrictions must be put on $\hat{H}$ and Ψ_i (ref. 4). Briefly, there must be a lower limit to the possible energy of the decayed nuclei, and for decay to take place, the spectrum of energies must contain a continuum. This is sufficient to show that the decay rate of any state must ultimately become slower than exponential at very large times. The mean energy of the initial state Ψ_i must be finite, which is sufficient to prove that the initial decay rate of any state must be slower than exponential for small times.

In fact the decay of an isolated quantum state can never be exponential. This is because the products of the decay at an earlier time can recombine to form the initial state later. (This can be seen by subdividing the interval t into two intervals t_1 and t_2 . The probability amplitude $A(t)$ must satisfy $A(t_1+t_2) = A(t_1)A(t_2) +$

$\langle \Psi_i | \exp(-i\hat{H}t_2) | \Psi_d(t_1) \rangle$, where Ψ_d is the state of the decay products at time t_1 . The second term is the 'memory' term that is additional in the quantum-mechanical treatment⁵, modifying the classical effect represented solely by the first term.) Thus quantum mechanics implies that the system has a memory, so that it is possible to determine the absolute age of a sample by examining its decay law, which is not possible for pure exponential decay.

In the new experiments, Norman *et al.*² searched for deviations from exponential



Rutherford included his celebrated exponential curves on his escutcheon.

decay for short and long times in the radioactivity of ^{60}Co and ^{56}Mn . ^{60}Co , which has a half-life of 5.271 yr, was made from ^{59}Co by neutron irradiation in a pulsed reactor. This defines a precise starting time for the experiment, and Norman *et al.* examined the decay law for the first 10^{-4} half-lives. Similarly, the decay of ^{56}Mn , with a half-life of 2.5785 h, was examined over 45 half-lives, by which time the initial activity had fallen by a factor of about 3×10^{14} . Neither isotope revealed any deviation from exponential decay, and nor has any other test⁶. This seems to be a paradox.

The resolution of this paradox lies in the form that Ψ_i must have if we are to speak of a decaying state. Essentially the decay time τ of Ψ_i must be long in comparison with the characteristic evolution times τ_w associated with the interactions which (nearly) bind the decay products of Ψ_i . Or equivalently, Ψ_i must be constructed from a wavepacket which is much narrower in energy than the range spanned by the interaction responsible for the decay⁷. Under these circumstances the memory time is of the order of τ_w , very much less than the decay time τ , and the memory term then becomes negligible.

Significant deviations from exponential decay occur actually only for very small⁸

and very large times. The intrinsic memory time τ_w , and the exponential decay time τ have already been identified. Two other times are important⁹: $\tau_E = \hbar/E_0$ where E_0 is the energy released in the decay and $\hbar$ is Planck's constant, and $\tau_L \sim 3\tau \log_e(E_0\tau/\hbar)$. One always has $\tau_w < \tau$. For decays with very small energy release, however, τ_E increases and τ_L can decrease into regions where observation is possible. It can be shown that exponential decay sets in only after the maximum of τ_E and τ_w and lasts until τ_L , so for decays near threshold ($E_0 \rightarrow 0$) there is no exponential decay region¹⁰.

But to see deviations from exponential decay, the decay products must be allowed to recombine. This is not a problem at short times; indeed a straightforward estimate of the time required for the decay products to separate by more than a de Broglie wavelength (the characteristic distance for quantum systems) gives τ_E , but for long times the situation is different. Any fluctuations, including those associated with observation in quantum mechanics, will destroy the recombination effect and therefore produce exponential decay. Furthermore, the decay constant will be τ , the undisturbed value, and independent of the characteristic timescale of the fluctuations if these occur predominantly in the exponential region of the decay⁹. This is likely to prevent the observation of post-exponential decay, particularly in nuclear systems, where τ_i is large. For ^{56}Mn , $\tau_L \approx 3$ weeks, and the decay products must be able to recombine after this time; $\tau_E \approx 10^{-21}$ s so the pre-exponential decay will also not be observed.

Thus it seems unlikely that nuclear decays will show deviations from the exponential-decay law which they made famous. The most promising process for observing non-exponential decay is the near-threshold photodetachment of electrons from negative ions using highly stabilized lasers¹¹. In this case, there is a chance of extending τ_E , the length of the pre-exponential region, and of shrinking τ_L , the time of onset of post-exponential decay. The correct combination of circumstances does not seem to arise naturally to produce deviations from exponential decay, which is why it is such an accurately fulfilled law, even out to 45 half-lives. □

1. Rutherford, E. *Sber. Akad. Wiss. Wien. Abt. 2A* **120**, 300 (1911).
2. Norman, E.B. *et al. Phys. Rev. Lett.* **60**, 2246–2249 (1988).
3. Fonda, L., Ghirardi, G.C. & Rimini, A. *Rep. Prog. Phys.* **41**, 587–631 (1978).
4. Khalifin, L.A. *JETP* **6**, 1053–1063 (1958).
5. Williams, D.N. *Commun. Math. Phys.* **21**, 314–333 (1971).
6. Nikolaev, N.N. *Soviet Phys. Usp.* **11**, 522 (1968).
7. Rice, O.K. *Phys. Rev.* **35**, 1538–1550; 1551–1558 (1930).
8. Chiu, C.B., Sudarshan, E.C.G. & Misa, B. *Phys. Rev. D* **16**, 520–529 (1977).
9. Greenland, P.T. & Lane, A.M. *Phys. Lett. A* **117**, 181 (1986).
10. Rzgżewski, K. *et al. J. Phys. B* **15**, L661–L667 (1982).
11. Haan, S.L. & Cooper, J.J. *Phys. B* **17**, 3481–3492 (1984).

P.T. Greenland is in the Theoretical Physics Division, Harwell Laboratory, Didcot OX11 0RA, UK.

Archaeology

Adding grist to the mill

Paul G. Bahn

MILLSTONES for grinding grain are fairly common in Roman settlements, particularly those around the Mediterranean. The attention of scholars has traditionally focused on the technological aspects of Roman millstones, and the study of these artefacts as evidence for trade networks was greatly neglected until a long-term investigation begun some time ago by David Peacock¹. The latest results from this programme are now reported by Olwen Williams-Thorpe and Richard Thorpe²⁻⁵ who use a combination of analytical techniques to pinpoint the precise source of the millstones. The results reveal intriguing details of trade patterns in the Mediterranean.

There are two principal types of millstones: large, hourglass-shaped mills powered by animals and called the Pompeian type because of their abundance at that site; and smaller, flatter, hand-driven cylindrical rotary mills called querns. They were made of various rocks, but igneous rocks were popular because they were rough enough to grind effectively, yet hard enough to prevent excessive grit becoming detached and entering the food.

By determining the rock type it has often been possible to pinpoint its geological source. The basic investigative technique is petrographic analysis, in which thin sections of the stones are examined under a microscope. This can be a reliable guide to source, but in some cases it is insufficiently accurate, with results pointing to several production areas.

The new work²⁻⁵ by Williams-Thorpe and Thorpe combines petrographic analysis and a more detailed determination of composition by wavelength dispersive X-ray fluorescence. This technique of geochemical analysis, requiring only one or two grams of rock crushed to a fine powder, is effective in pinpointing a precise source.

So far, the investigators have examined millstones and possible sources in various regions of the Mediterranean and Europe and have obtained some intriguing results. In southern Britain, for example, there are three known donkey-mills of Pompeian type, all of grey lava. They were tested against sources in the Eifel area of Germany and in France's Massif Central, as these were the two centres of known lava millstone production closest to Britain. The Romano-British specimens from London and from Corfe Mullen (Dorset) turn out to be from two different sources in the Chaîne des Puys, Central France, thus demonstrating the existence of millstone production in at least two localities

in that area². This result confirms Peacock's earlier identification of the source of the London specimen through thin-section analysis¹. So the German sources must have supplied Britain only with the smaller, cylindrical mills. The third British donkey-mill, from Hamworthy (Dorset), came from Sardinia but is probably post-Roman.

The situation in the Mediterranean is far more complex: here, a total of 113 millstones sampled, dating from the third century BC to the sixth AD, were found to come from 16 different areas³. The production and trading patterns vary with the raw material. Grey lava millstones



Fig. 1 Rotary millstones from the shipwreck of Isla Pedrosa, Spain (from ref. 5). Photograph courtesy of O. Williams-Thorpe/Museo Arqueologico Provincial di Girona, Spain.

seem to come from four main regions: Morocco; north-east Spain/south-east France; southern Italy/Sicily/north Libya/Tunisia; and Greece/the Aegean. Within each area, the stones were traded up to 820 km from their source, but there was almost no trade in this material between the four areas as each had adequate sources of its own.

Superimposed on this basic pattern of restricted use, however, was a trading network that cut across the boundaries, spanning greater distances, and involving rhyolite mills from Sardinia and leucite mills from Italy: these were transported up to 1,350 and 1,500 km, respectively, and were clearly more highly prized than the local lavas, perhaps because they were of better quality. The Italian leucite was traded widely, but apparently was not exported to Sardinia, where rhyolite was obtainable.

Although some millstones were transported overland (up to 250 km in North Africa, up to 125 km in Sardinia), most were probably moved by sea. They could have been carried on grain-collecting ships, which headed west and south from

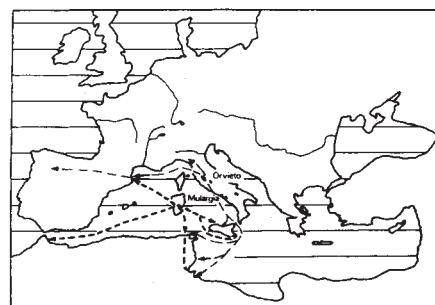


Fig. 2 Sources of, and distances travelled by, the better quality stones — rhyolite from Mulargia, Sardinia, and leucite from Orvieto, Italy. (From ref. 3.)

Italy to North Africa, Spain, Sardinia and Sicily, the 'bread-baskets' of the Roman empire. Rome required huge quantities of grain, and this shipping was organized on a grand scale. The millstones would have served not only as trade objects but also as ballast in the ships on the outward voyage.

The islands were natural stopping places on these trips, a fact which would explain how millstones from their sources reached North Africa but not vice versa. Sardinia, for instance, was an important Roman settlement and trade centre from the third century BC till AD 500. Itself an important contributor of grain, it also lay on a shipping route from Italy to North Africa and had excellent harbours. Its rhyolite stones were widely traded in the western Mediterranean, being found in Spain, Morocco, Tunisia and Sicily⁴. On the return voyage to Italy the ships were full of grain, with no room for millstones; this explains why so few specimens have been found in Italy, which in any case had its own high-quality sources.

Some stones were transported unfinished, including at least one of the 59 millstones found in the cargo of a wrecked vessel at Isla Pedrosa, off the coast of north-east Spain near the settlement of Ampurias (Fig. 1). The ship contained Campanian amphorae, dating to BC 150 or 140, and was probably Italian. Williams-Thorpe and Thorpe analysed twelve of its millstones, and as ten of these were of local manufacture (from the lava of the Olot/Girona area) but the two others came from Agde (southern France) and Sicily, they believe that either the vessel had called at Agde and Sicily on its way from Italy or, more likely, that all the stones were picked up from Ampurias. If the latter explanation is correct, then this cargo is evidence for a secondary trade in millstones. Certainly Ampurias was an important trading centre, as is reflected in its millstones from five sources; similarly, the great trading city of Carthage in Tunisia had links with nine sources (Fig. 2).

Only a little information is provided in classical texts about millstone production, and these mention a mere handful of sources like Etna. For example, there is no

documentary reference to Sardinia or to leucite sources. Historical and archaeological evidence add to this picture, and quarries have been discovered in Italy, France and Germany. More field-work is needed to identify quarry sites elsewhere from their rough-outs and broken or unfinished stones. Clearly, however, petrographic and geochemical analyses have succeeded, in only a few years, in revolutionizing our knowledge of

millstone production and trade in the Roman world, and no doubt a fuller pattern will emerge as these analyses are extended to all parts of the empire. □

1. Peacock, D.P.S. *World Archaeol.* 12, 43 (1980).
2. Williams-Thorpe, O. & Thorpe, R.S. *Archaeometry* 30, 273-287 (1988).
3. Williams-Thorpe, O. *J. Archaeol. Sci.* (in the press).
4. Williams-Thorpe, O. & Thorpe, R.S. *Oxford J. Archaeol.* (in the press).
5. Williams-Thorpe, O. & Thorpe, R.S. *Vitrina* 2, 49 (1987).

Paul G. Bahn is an archaeology writer at 428 Anlaby Road, Hull HU3 6QP, UK.

Luis Walter Alvarez (1911-1988)

ON 1 September 1988, Luis Walter Alvarez, one of the most creative scientists of our time, died at the age of 77, a year after surgeries that first partially removed a benign brain tumour and then exised a cancer from his oesophagus. Alvarez had three simultaneous successful careers — as a scientist, as an inventor, and in aviation. He won the 1968 Nobel prize for physics for his development and use of the hydrogen bubble chamber to discover a number of particles. In aviation, Alvarez was awarded the 1946 Collier air trophy for his radar-based blind landing system which was first used in the Battle of Britain. In 1978, he was inducted into the Inventors' Hall of Fame for his clever physics instruments, brilliant aviation radar systems and ingenious optical devices commercialized in two successful companies.

Recitation of Alvarez's honours provides but a pallid profile of this colourful scientific swashbuckler. A third-generation Californian, the son and grandson of physicians, Luie, as he was known to all, grew up in Rochester, Minnesota. As an undergraduate at the University of Chicago, Luie's life-long fascination with optics began and was the subject of his first paper, a measurement of the wavelength of light using a phonograph record under the parlour lamp.

He then constructed and operated one of the first Geiger counters in the United States, despite mistakes like reversed d.c. polarity and oxygen-contaminated filling gas. Here was a characteristic of Luie's later work — luck. Arthur Compton suggested they collaborate, and Luie's first summer was spent on the roof of the Geneva Hotel in Mexico City with his Geiger telescope. The discovery of the east-west effect was published with Alvarez as first author. The generosity towards his young colleague that Compton demonstrated became a hallmark of Luie's relationships with junior members of his later enterprises.

After obtaining his PhD, Luie went to the University of California radiation laboratory to become a nuclear physicist. Hans Bethe's three-part compendium of nuclear physics guided Luie to choose measurements Bethe said could not be done or that would disprove a Bethe

assertion. In just 4 years, Luie discovered that nuclei cannibalize their atomic electrons, tritium radioactivity and helium-3 stability (with Bob Cornog), and the neutron's magnetic moment (with Felix Bloch). He demonstrated the spin dependence of the nuclear force (with Ken Pitzer), a new standard of length (with Jake Wiens), and heavy-ion collisions.

The war in Europe and the visit of the Tizard mission ended Luie's nuclear physics career. Soon he was in Boston applying high-frequency radio waves to

then invented the so-called tandem van de Graaff and with the Soviet atomic-bomb explosion, Luie became involved with the improbable behemoth MTA which was to breed fissile material for weapons.

In 1953, Luie heard young Don Glasser describe experiments in which cosmic rays caused boiling of test-tube quantities of ether. Luie immediately envisioned a gigantic pot of liquid hydrogen operating like a Wilson cloud chamber. This extrapolation made even Ernest Lawrence, the founder of Lawrence Berkeley laboratory, blanch, but displayed Luie's intellectual self-confidence. The resulting physics and technology led to his Nobel prize.

After a dozen years of exploiting his bubble chambers and further establishing 'big science', Luie abandoned both. He began pioneering superconducting magnets for cosmic-ray balloon studies and to search for the magnetic monopole. His interests now moving towards astrophysics, he searched for the undiscovered chambers in the Chephren pyramid.

In 1979, recently retired from the Uni-

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Luis Alvarez — one of the most creative scientists of our time.

military goals. There Luie invented the linear phased array for EGALE, the first radar bombing system. He also invented VIXEN, a system to outfox German submarines. Luie solved the problem of landing planes in bad weather by inventing ground controlled approach (GCA). Finally, "Alvarez's Folly", an early-warning radar system (MEW), set the standard for the post-war US air-defence system.

After a stint with Fermi in Chicago, Alvarez went to Los Alamos and invented the exploding wire technique that provided the implosion of fissile bomb material. He then developed pressure gauges which he personally used on the Trinity test and Hiroshima mission.

Returning to Berkeley after the war, Luie became a particle accelerator physicist. He had already invented the electron microtron, so the proton linear accelerator was his first post-war machine. He

versity of California, Luie made what might turn out to be his most important discovery. His geologist son Walter had collected 65-million-year-old clay which, when analysed by nuclear chemists Frank Asaro and Helen Michel, showed a high concentration of iridium. The abundance of this element was more characteristic of comets and asteroids than of material from the Earth's crust. The age of the clay coincided with the Cretaceous/Tertiary event in which dinosaurs and many other life forms were wiped out. Luie invented various models using these data, all but one of which he later discredited. His asteroid extinction theory emerged and the data that support it continue to grow.

W. Peter Trower

W. Peter Trower is at Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061, USA.

Molecular neurobiology

Separating receptor subtypes from their shadows

Eric A. Barnard

SUBTYPES of receptors were first identified by their pharmacological differences, either in their response to drugs *in vivo* or in their behaviour when binding ligands *in vitro*. But are these differences of real significance for receptor action? New data from many laboratories on various G-protein-linked receptors are helping to elucidate this question using molecular biology techniques. Two such papers in *Nature*, one by Capon and collaborators reported last month¹ and one by Numa and colleagues on page 355 of this issue², characterize subtypes of the muscarinic receptor. Lefkowitz, Caron and collaborators, on page 358 of this issue³, reveal a similarity between a β -adrenergic receptor and a 5-HT (serotonin) receptor. These reports, together with other recent data, help to explain the great diversity of subtypes of G-protein-linked receptors.

The muscarinic receptor has been classified into pharmacological subtypes termed M_1 and M_2 . The antagonist pirenzepine and the agonist McN-A-343 act with high physiological sensitivity or binding affinity on the M_1 subtype, found in mammalian brain cortex or hippocampus. These two drugs are much less effective on the M_2 subtype, which can be divided into ' M_2 cardiac' and ' M_2 glandular' subtypes by their relative affinities for the antagonists hexahydrostiladifenidol or AF-DX 116^{4,5}.

Equivalent diagnostic drugs have been used to define the subtypes of numerous other G-protein-linked receptors. But how significant are these categories? Sometimes they are inconsistent between species (the receptor in chick heart, for example, does not fit the M_2 cardiac profile⁶) and often the affinity differences used as criteria for receptor subtypes are not very striking. Could the multiple subtypes which are found result from different modifications or from different environments of the same receptor protein, or from a receptor which may be free or exist in various G-protein complexes?

For a long time these basic questions concerning receptor subtypes remained completely open. More recently, however, important differences in sequence have been found in the cloned DNAs encoding a given receptor type. Thus, five different genes have so far been identified that encode muscarinic receptors and there are indications of yet more to come, as a paper by Bonner *et al.* in a recent issue⁷ of *Neuron* demonstrates. Receptor isoforms, in the true structural sense, do exist and are increasingly found to be

more numerous than would have been expected from the known pharmacology of that receptor. We have recently shown⁸ that there are at least three isoforms of the GABA_A receptor α -subunit, functionally distinct, although only two subtypes are well-recognized pharmacologically. This is yet another example of greater multiplicity at the DNA level.

These results allow a more precise question about receptor subtypes to be asked. Do isoforms of a receptor correspond to functionally distinct subtypes, and vice versa? To answer this, receptor functions need to be defined in specific molecular terms: as well as the binding parameters, the particular transduction systems and channel couplings used *in vivo* are clearly relevant. The question can be addressed using the recently developed technologies for expression of recombinant receptors, via messenger RNA translation in the *Xenopus* oocyte or via DNA transfection into cell lines lacking the receptor in question. These approaches permit functional assays at all the levels mentioned above.

The papers from Capon's group¹ and Numa's group² go some way to answering the question. But unfortunately the nomenclature for the muscarinic receptor subtypes used by the two groups differs; indeed, the terminology used in different laboratories can sometimes be contradictory. Now that subtypes are being established on the basis of the DNA encoding them, cannot workers on muscarinic receptors agree on a universal notation? In the meantime, to try to remove confusion for non-specialists, the table shows the relations between the terminologies currently used. In this article I use M1–M5 to describe the subtypes, assigning these as in ref. 1.

Peralta *et al.*¹ expressed recombinant human M1–M4 isoforms by stable transfection in a kidney cell line. They showed

that M2 and M3 act efficiently in inhibiting adenylyl cyclase but not in activating phosphoinositide (PI) turnover, whereas the reverse is true for M1 and M4. The effect of M1 in mediating PI turnover is in agreement with the results of Lai *et al.*⁹ who obtained, in a transfected fibroblast cell line, stable expression of the rat M1 subtype gene and found that it introduces agonist-dependent PI hydrolysis, inhibited strongly by pirenzepine and only weakly by AF-DX 116. Fukuda *et al.* in this issue² report the results of transfecting hybridoma-glioma NG108-15 cells with DNAs encoding the porcine or rat M1–M4 receptors. In addition to showing differential effects on adenylyl cyclase (M2 and M3) and PI turnover (M1 and M4) comparable with those of Peralta *et al.*¹, they show that characteristic ion-channel responses to muscarinic agonists can be introduced by the transfected receptor subtypes. Only those receptors strongly coupled to PI turnover (M1 and M4) display agonist-induced opening of a Ca^{2+} -dependent K^+ channel and agonist-dependent inhibition of the voltage-sensitive K^+ channel which is responsible for the M-current⁷.

The receptor classified by pharmacologists as M_2 (glandular) is intriguing. Numa and collaborators find¹⁰ that the cloned porcine M4 (intronless) gene produces, upon injection of its RNA into *Xenopus* oocytes and analysis by binding of the three diagnostic ligands noted initially above, a muscarinic receptor with M_2 (glandular) pharmacology. This effect of M4 was distinguished thus from the M1, M2 and M3 subtypes expressed in the oocyte in parallel. But Bonner *et al.*⁷ have expressed a newly found human or rat muscarinic receptor gene, M5, in CHO cells and find first, that the affinities for pirenzepine and AF-DX 116 are closest to those of the M_2 (glandular) type. (But the ' μM ' values in their Table I should read ' nM ' for all the pirenzepine values because of a printing error, as the authors confirm.) Second, they find that M5 is coupled to PI turnover and not to adenylyl cyclase. However, no appreciable M5 mRNA is detected in the exocrine glands tested, nor in any of the other tissues⁷. As M4 mRNA is the main form present in the pancreatic¹¹

Subtypes of muscarinic receptors and their transductions

As defined in the reports of:			Pharmacological equivalent*	Primary transducing effect:		Channel couplings
ref. 1	ref. 2	ref. 7		PI turnover	Cyclase inhibition	
M1	mAChR I	m1	M_1	Yes ^{1,2,9}	No ¹	O ^{2,17}
M2	mAChR II	m2	M_2 (cardiac)†	No ^{1,2}	Yes ¹	
M3	mAChR IV	m4	?	No ^{1,2}	Yes ¹	
M4	mAChR III	m3	M_2 (glandular)	Yes ^{1,2}	No ¹	O ²
M5		m5	?	Yes ⁷	No ⁷	ND

O, Opening of a Ca^{2+} -dependent K^+ -channel and closing of the M-current K^+ channel; ND, not done. Nomenclature of ref. 1 is used in this article

* For the mammalian tissues so far analysed with the discriminatory set of drugs.

† Also identified by finding M1 or M2 protein sequence in peptides from the receptor purified from, respectively, porcine cortex or porcine heart (see ref. 11).

and salivary (Bonner, T. I. *et al.*, submitted) glands, and as the muscarinic PI effect is characteristic of these glands (see ref. 11), M4 again is thought likely to be M₂ glandular. Identification of M3 and M5 is required, however. The five muscarinic subtypes so far found by DNA cloning fall into two clear functional groups (see table). More subtle functional or regulatory differences should now be sought, to account for the individuality of the isoforms (which are highly conserved — there is 89–96 per cent amino-acid identity between rat, pig and man^{7,10} — as are the GABA_A receptor isoforms).

In their paper this issue³, Caron, Lefkowitz and colleagues report an interesting case of the cloning of a DNA for a G-protein-linked receptor by cross-receptor-hybridization. The authors used transfection in COS cells to identify a human genomic clone which they had obtained by hybridization at reduced stringency with a gene probe for the β_2 -adrenergic receptor. They obtain a ligand-binding profile clearly characteristic of a serotonin receptor (5-HT_{1A} type) in the transfected cells, and confirm this result immunochemically. The amino-acid identity between this receptor and any adrenergic receptor in the seven presumed transmembrane domains found in all of them is about 45 per cent. The level of nucleotide homology in very restricted regions is obviously sufficient to allow the low-stringency hybridization of the β_2 and the 5-HT_{1A} DNAs, providing hope that genes encoding other subtypes in this series may be cloned using this technique.

This report of Fargin *et al.*³ provides a subtype sequence which can be compared with another revealed in a notable recent achievement by Julius *et al.*¹². These authors, again, used the oocyte receptor-expression system, in this case cloning in an RNA expression vector and screening the RNAs for induction of a 5-HT response in oocytes, thus selecting pools sequentially until a single positive clone was isolated. The exceptional sensitivity of the 5-HT_{1C} response in the oocyte (down to 5 fg receptor mRNA) and a rich source of receptor mRNA (choroid plexus) make this an especially favourable case for this strategy, previously initiated for the same receptor by Lubbert *et al.*¹³. Both the 5-HT_{1C} and the 5-HT_{1A} receptors^{3,12} have the seven membrane-crossing predicted structure now familiar for G-protein-linked receptors. If the two are aligned on this basis, their similarity (20 per cent overall) is low (suggesting different functions), indeed less than that of either with the human β -adrenergic receptor. Of those sites where identical residues do occur, 70 per cent are in the seven membrane domains but even when only those segments are considered, identity is only 35 per cent. Hence, these two 5-HT₁ receptor subtypes differ even more

than any of the pairs of muscarinic subtypes⁷. The 5-HT_{1C} receptor has been linked to PI turnover, although the transduction system of the 5-HT_{1A} receptor is still uncertain.

These and other recent examples of receptor functional subtypes based upon different DNA sequences establish this situation as a general phenomenon. Why are the subtypes maintained with great evolutionary stability? One reason must be for tissue and regional specialization. Sometimes this is obvious, as in the case of cardiac muscle versus brain muscarinic subtypes. But all the M1–M4 genes are expressed to a similar extent in the brain¹¹. The subtypes of a receptor can confer different sensitivities to the transmitter and to possible modulators, as well as alternative transduction pathways for different effector systems^{1,2,8}. Pre- versus postsynaptic locations of a given receptor often require a different behaviour and hence different subtypes. Finally, some of the isoforms, especially those seen by genomic cloning, are likely to be developmental specializations. How receptor subtypes may operate here is beautifully illustrated by the recent report¹⁴ of the late Stephen Schuetz and colleagues. The nicotinic receptor in mammalian skeletal muscle has embryonic (γ) and adult (ϵ) subtypes with a clear functional difference¹⁵, ϵ producing a channel of fourfold shorter lifetime and twice the conductance of the γ -channel. Jaramillo *et al.*¹⁴ deduce that it is the properties of the γ -channel which permit the current flow needed for early spontaneous contractions. These contractions will lead to the darwinian selection¹⁶ of the final innervation and initiate neuromuscular differentiation, which then will require the adult receptor subtype. Presumably, equivalent mechanisms occur in other receptors in the developing brain.

As these and other clues to the meaning of receptor subtype diversity are uncovered, we can recall what St Thomas Aquinas said about types: "Types and shadows have their ending, for the newer rite is here". To distinguish the receptor types from the shadows, the newer rite is molecular biology. □

1. Peralta, E. G. *et al.* *Nature* **334**, 434–437 (1988).
2. Fukuda, K. *et al.* *Nature* **335**, 355–358 (1988).
3. Fargin, A. *et al.* *Nature* **335**, 358–360 (1988).
4. Birdsall, N. J. M. *et al.* *Biochem. Soc. Symp.* **52**, 23 (1986).
5. Ladinsky, H. *et al.* *Trends pharmac. Sci. Suppl.* **9**, 44 (1988).
6. Brown, J. H., Goldstein, D. D. & Masters, S. B. *Molec. Pharmac.* **27**, 525–531 (1987).
7. Bonner, T. I. *et al.* *Neuron* **1**, 403–410 (1988).
8. Levitan, E. S. *et al.* *Nature* **335**, 76–79 (1988).
9. Lai, J. *et al.* *Life Sci.* **42**, 2489–2502 (1988).
10. Akiba, I. *et al.* *FEBS Lett.* (in the press).
11. Peralta, E. G. *et al.* *EMBO J.* **6**, 3923–3929 (1987).
12. Julius, D. *et al.* *Science* **241**, 558–564 (1988).
13. Lubbert, H. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **84**, 4332 (1987).
14. Jaramillo, F. *et al.* *Nature* **335**, 66–68 (1987).
15. Mishina, M. *et al.* *Nature* **321**, 406–411 (1986).
16. Edelman, G. M. *Neuronal Darwinism* (Basic, New York, 1987).
17. Jones, S. V. P. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **85**, 4056–4060 (1988).

Eric A. Barnard is director of the MRC Molecular Neurobiology Unit, Cambridge CB2 2QH, UK.

Daedalus

Cleaning up the papers

THE computer revolution once promised to give us the "paperless office". Instead, it has created even more paper: print-outs, document drafts, reports and memos, all in embarrassing profusion, all losing importance rapidly as they are replaced by updated versions, but gaining nuisance value from the proliferation of their numberless photocopies. Daedalus is resigned to the bureaucracy; but he is distressed at the great waste of resources. So he is inventing an erasable and re-usable paper.

His pilot scheme borrows from the chemistry of picture restoration. The white lead-carbonate pigment in an old oil painting is slowly attacked by atmospheric hydrogen sulphide, and darkens over the years into black lead sulphide. The picture restorer applies hydrogen peroxide to bleach this back to white lead sulphate, restoring the original brightness and sparkle of the painting. Many types of paper carry a loading of white mineral material, so DREADCO's chemists are devising a paper loaded with lead sulphate. A special dot-matrix printer, its ribbon impregnated with sodium sulphide, locally reacts the sulphate to black lead-sulphide characters. A complementary erase head, generating ozone from an electric discharge, can oxidize unwanted characters back to lead sulphate again. Thus the draft can be edited *in situ*, or even erased completely back to pristine white paper. The whole system is rather poisonous and smelly, but Daedalus hopes to invent a less noxious chemistry and a suitably durable base material when he has proved the basic concept.

Once it is commercially perfected, DREADCO's Repaper® will blissfully simplify office practice. Old documents and reports will be run through the printer again and updated to new ones; vast reams of bug-ridden or outdated computer tabulations will be effortlessly recycled; telexes and fax machines will record new transient messages while erasing old ones. In every office, the waste-paper basket and the shredder will go hungry. Old computer print-out, the wide, free scribbling pad of a whole generation, will suddenly vanish from playrooms and studies. Paper will become *valuable* again, retrieved and hoarded in the informal office economy like paper clips and floppy disks.

But every invention has its drawbacks. Organizations that unthinkingly use Repaper for cheques, airline tickets, security passes and certificates of competence will soon discover its true potential. And in love-letters, frank confessions, and statistical reports, Daedalus' new medium will certainly convey its own strong unstated message.

David Jones

Similarity of endothelin to snake venom toxin

SIR—We wish to point out the striking functional and structural similarities between porcine endothelin¹ and certain newly isolated toxins from the burrowing asp, *Atractaspis engaddensis*². These sarafotoxins S6 cause cardiac arrest when injected into mice ($LD_{50} = 0.01\text{--}0.2 \mu\text{g}$ per gram body weight), probably as a direct result of coronary vasospasm³. Similarly, endothelin is reported to be one of the most potent vasoconstrictors known.

The amino-acid sequences of these peptides are remarkably similar (see figure). Both peptides consist of a single

ENDOTHELIN	C	S	C	S	S	L	M	D	K	E	C	V	Y	F	C	H	L	D	I	I	W
SARAFOTOXIN S6b	C	S	C	K	D	M	T	D	K	E	C	L	Y	F	C	H	Q	D	V	I	W

Amino-acid sequences of endothelin and sarafotoxin S6b. One-letter amino-acid code. Boxes sequence identity. Sequence of endothelin from ref. 1 and sequence of sarafotoxin S6b from ref. 2.

21-residue peptide chain containing four cysteines. All four cysteinyl residues appear in identical positions, and the overall sequence homology between endothelin and sarafotoxin S6b is 67 per cent. The peptides almost certainly share a common ancestor.

Although the role of endothelin in the porcine cardiovascular system remains to be determined, sarafotoxins S6 function as offensive weapons to rapidly incapacitate prey species. The almost simultaneous discovery of these peptides in such very different circumstances underscores the versatility of nature, and serves as yet another example of the conversion by an organism of a biologically active peptide into a toxic principle.

C. Y. LEE

Department of Pharmacology,
College of Medicine,
National Taiwan University,
Taipei, Taiwan, ROC

VINCENT A. CHIAPPINELLI
Department of Pharmacology,
St Louis University School of Medicine,
St Louis,
Missouri 63104, USA

SIR—We have found an intriguing homology between endothelin (ET), the vasoconstrictor peptide we recently identified^{1,4} in the mammalian vascular endothelium, and a group of peptide toxins (sarafotoxins S6) recently purified and sequenced² from the venom of Israeli burrowing asp *Atractaspis engaddensis*. All these peptides consist of 21 amino-acid residues with two pairs of half-cystine residues in identical positions and a hydrophobic tail with a carboxy-terminal tryptophan residue (see figure). The strong sequence similarities (52–67 per cent identities or 71–81 per cent identities including conser-

vative substitutions) convincingly suggest that the toxins and ET have a common evolutionary origin.

These peptides resemble each other not only in structure but also in the biological activities. Sarafotoxins S6, like ET, have a potent coronary constrictor activity³; their strong cardiac toxicity is chiefly the result of a persistent coronary vasoconstriction. When administered intravenously to mice in doses as low as $10\text{--}200 \mu\text{g kg}^{-1}$, sarafotoxins S6 cause a severe coronary vasospasm that, accompanied by electrocardiographic changes similar to those seen in severe acute myocardial infarction, eventually leads to cardiac arrest and death^{2,5}. We have found that intravenous injection of suprapressor doses of ET is also highly lethal to rats; injection of $10\text{--}20 \mu\text{g kg}^{-1}$ usually kills an animal by severe coronary insufficiency (unpublished observations). The extremely long-lasting vasoconstrictor activity of ET, which is quite difficult to wash out¹, also exemplifies the toxin-like properties of this mammalian peptide.

The existence of cognate peptides of ET in the exocrine venomous glands of snakes suggests that ET, or related peptides, are expressed in mammalian organs other than vascular endothelium. Perhaps a

ET	C	S	C	S	S	L	M	D	K	E	C	V	Y	F	C	H	L	D	I	I	W
S6 b	C	S	C	K	D	M	T	D	K	E	C	L	Y	F	C	H	Q	D	V	I	W
S6 a1	C	S	C	K	D	M	T/S	D	K	E	C	L	Y	F	C	H	Q	D	V	I	W
S6 c	C	T	C	K	D	M	T	D	E	E	C	L	N	F	C	H	Q	D	V	I	W
	1					5					10					15					20

Alignment of porcine/human endothelin (ET) and several isoforms of sarafotoxins S6 from venom of *A. engaddensis*. Identical and conservatively substituted amino-acid residues enclosed in boxes; half-cystine residues in bold.

number of distinct peptides of the 'ET/sarafotoxin family' will be found in various tissues from a wide range of species. Chemical identification of such peptides would provide an excellent opportunity to investigate the structure-activity relationship as well as the physiological and/or pathophysiological significance of endothelin.

CHIKAHISA TAKASAKI

Department of Chemistry,
Tohoku University,
Sendai 980, Japan

MASASHI YANAGISAWA
SADAO KIMURA
KATSUTOSHI GOTO
TOMOH MASAKI

Institute of Basic Medical Sciences,
University of Tsukuba,
Tsukuba,
Ibaraki 305, Japan

1. Yanagisawa, M. *et al.* *Nature* **332**, 411–415 (1988).
2. Takasaki, C. *et al.* *Toxicon* **26**, 543–548 (1988).
3. Lee, S.-Y., Lee, C.-Y., Chen, Y.-M. & Koehva, E. *Toxicon* **24**, 285 (1986).
4. Itoh, Y. *et al.* *FEBS Lett.* **231**, 440–444 (1988).
5. Wollberg, Z. *et al.* *Toxicon* **26**, 525 (1988).

Negative or positive cloud optical depth feedback?

SIR—Roeckner *et al.*¹ have reported results from a pair of general circulation model (GCM) simulations — a present-day control and a perturbation with the solar constant increased by 2 per cent — in which the cloud optical depth could vary through its dependence on the prognostically calculated cloud liquid water. This simulation pair predicted a 3.3°C warming of the globally averaged surface-air temperature as a result of the 2 per cent solar constant increase.

To determine the contribution of cloud optical depth feedback to this warming, Roeckner *et al.* calculated the changes in the net radiative fluxes at the tropopause (top of their GCM) and at the Earth's surface, ΔN_t and ΔN_s respectively, that result when the equilibrium climate of the control simulation is modified by replacing the cloud liquid water with its value from the 2 per cent solar constant experiment. The results show that ΔN_t and ΔN_s are then of different sign, ΔN_t being positive and ΔN_s negative. Roeckner *et al.* conclude that "The net effect of clouds is to provide a negative feedback on surface temperature, rather than the positive feedback found in earlier general circulation model studies without considering cloud optical depth feedbacks". But whether or not the cloud optical depth feedback in these simulations is negative depends upon whether the change in the energy balance at the surface, characterized by ΔN_s , dominates the change in the energy balance for the planet, characterized by ΔN_t . If the cooling tendency from the decrease in ΔN_s is greater than the warming tendency from the increase in ΔN_t , the cloud optical depth feedback is negative. If the opposite occurs, the cloud optical depth feedback is positive.

To determine which of these two outcomes is more probable I have performed two control/perturbation simulation pairs with the radiative-convective model (RCM)^{2,3}. Other feedbacks included in this model for the present study are the positive feedbacks from water vapour (fixed relative humidity), cloud attitude (fixed cloud-top temperature, 228 K) and surface albedo, and where indicated, negative feedback from moist-adiabatic lapse-rate adjustment. For the control simulation the optical depth τ and emissivity ϵ of the RCM's single cloud-layer were fixed at $\tau = 1.11$ and $\epsilon = 0.62$. These values were also used for the perturbation simulation without cloud optical depth feedback. For the perturbation simulation with cloud optical depth feedback, τ and ϵ were changed by $\Delta\tau$ and $\Delta\epsilon$, with the values chosen to reproduce the ΔN_t and ΔN_s values for the total cloud liquid-water effect in ref. 1. The

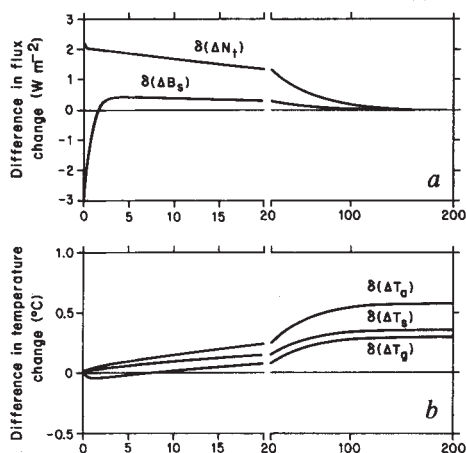
RCM study of cloud optical depth feedback

$\Delta\tau^*$	$\Delta\epsilon^*$	ΔN_1 (W m^{-2})	ΔN_s (W m^{-2})	2% S_0 ΔT_s^* ($^{\circ}\text{C}$)	2 $\times\text{CO}_2$ ΔT_s^* ($^{\circ}\text{C}$)
0.41(0.65)	0.16(0.28)	2.4	-2.9	5.3(2.6)	5.3(2.8)
0	0	0	0	4.2(2.2)	4.4(2.5)

*For a fixed lapse-rate of $6.5^{\circ}\text{C km}^{-1}$ and for a variable moist-adiabatic lapse-rate (in parentheses).

results of the RCM study are presented in the table for two perturbations, a 2 per cent solar constant increase and a CO_2 doubling, both for a fixed lapse-rate of $6.5^{\circ}\text{C km}^{-1}$ and for a moist-adiabatic lapse-rate (shown in parentheses) which varies with temperature. The change in the surface air temperature ΔT_s for both perturbations is positive, and ΔT_s with cloud optical depth feedback ($\Delta\tau \neq 0$ and $\Delta\epsilon \neq 0$) is larger than ΔT_s without ($\Delta\tau = 0$ and $\Delta\epsilon = 0$). Consequently, this RCM analysis indicates that the cloud optical depth feedback is positive.

To show how this positive cloud optical depth feedback occurs, the time evolution of the RCM solutions for the solar-con-



stant perturbation with the moist-adiabatic lapse-rate is displayed in the figure in terms of $\delta(\Delta\psi) = [\Delta\psi$ (with cloud optical depth feedback) - $\Delta\psi$ (without cloud optical depth feedback)], where $\Delta\psi = [\psi$ (perturbation) - ψ (control)] and ψ is any quantity. The surface of the RCM was taken to be ocean of 1 m depth. The timescale of the response of the RCM is related to this depth, and therefore has no physical significance. Initially the change in the net radiation at the tropopause, $\delta(\Delta N_1)$, is positive and the change in the net surface flux, $\delta(\Delta B_s)$, is negative. Here $B_s = N_s - H_s$, where H_s is the combined surface sensible and latent heat fluxes. As time increases both $\delta(\Delta N_1)$ and $\delta(\Delta B_s)$ approach zero as a result of the Earth-atmosphere system approaching equilibrium in each simulation. But

$\delta(\Delta B_s)$ does not approach zero monotonically from its initially negative value, but instead first becomes positive and then approaches zero monotonically. As a result, the change in the surface temperature $\delta(\Delta T_s)$ is negative initially but then becomes positive following the change in sign of $\delta(\Delta B_s)$.

Finally, I have investigated the fundamental assumption of Roeckner *et al.* that the change in sign of the surface temperature $\delta(\Delta T_s)$ is determined by the change in the net radiation at the surface $\delta(\Delta N_s)$ and is independent of the change in the net radiation at the top of the atmosphere $\delta(\Delta N_1)$. By varying $\Delta\tau$ and $\Delta\epsilon$ independently in the RCM for both the fixed and moist-adiabatic lapse-rate cases, we find that both the sign and magnitude of the temperature changes are determined by the change in the net radiation at the top of the atmosphere, $\delta(\Delta N_1)$, and are virtually independent of the change in the net radiation at the surface, $\delta(\Delta N_s)$, which is quite the opposite to the fundamental assumption made by Roeckner *et al.*

In conclusion, my RCM study indicates that the cloud optical depth feedback in the Roeckner *et al.* GCM simulations is positive. Of course the RCM is not the GCM, so the final, conclusive result will have to await a comparison of the appropriate GCM simulation with that in ref. 1. In any event, because of the complexity of cloud processes in nature, many additional modelling and observational studies will be required to determine the sign of the actual cloud optical depth feedback, let alone its magnitude.

MICHAEL E. SCHLESINGER

Department of Atmospheric Sciences and
Climatic Research Institute,
Oregon State University,
Corvallis,
Oregon 97331, USA

ROECKNER REPLIES—In response to Schlesinger's RCM feedback analysis of our GCM results, we performed another pair of GCM control/perturbation experiments, with cloud optical properties prescribed identically in both experiments using zonal/time averages as a function of height and latitude from the control experiment, in which the optical properties of the clouds were allowed to vary according to their dependence on the prognostically determined liquid water content¹.

The new pair of equilibrium-climate states serves as a reference for estimating the feedback that occurred in the earlier pair of experiments with variable cloud optics¹. As in Schlesinger's RCM analysis, the cloud optical depth feedback can now be evaluated directly as the difference δ of two equilibrium temperature changes ΔT that result from increasing the solar constant by 2 per cent in GCM simulations with and without the respective feedback.

Our results confirm Schlesinger's conclusion that the cloud optical-depth feedback in our GCM is positive, in contrast to our previous conclusion¹ and to earlier estimates⁴. The global and annual mean feedback $\delta(\Delta T)$ is largest in the upper troposphere (1.1°C) but is still positive at the surface (0.3°C) despite a negative net surface-radiation response. The chain of events creating the positive cloud optical depth feedback may be summarized as follows. The externally induced warming (solar constant increase by 2 per cent) causes a liquid-water increase in all types of cloud (but predominantly in high (cold) clouds) in accord with thermodynamic theory⁵. The associated increase of high-cloud infrared emissivity by approximately 15 per cent dominates the net planetary radiation change, that is, more long-wavelength radiation is trapped within the atmosphere than solar radiation is screened from the surface. The resultant upper tropospheric heating influences the surface by reduced convective overturning and less evaporative cooling at the surface. Thus, as Schlesinger infers, the cloud-induced negative surface-radiation change is not balanced by surface cooling¹ but instead by a decrease in the latent-heat flux at the surface.

Although our present GCM experiments cannot provide a final answer to the question of cloud feedback in the real world, they show that the negative cloud optical-depth feedback expected for optically thick clouds^{1,4} may be superseded by the positive feedback resulting from the emissivity change of optically thin cirrus clouds. Accordingly, the sign of the feedback may be crucially dependent on the spatial distribution of cloud amount and cloud optical properties in the control climate state.

ERICH ROECKNER

Meteorological Institute,
Hamburg University,
D-2000 Hamburg 13, FRG

1. Roeckner, E., Schlese, U., Biercamp, J. & Loewe, P. *Nature* **329**, 138–140 (1987).
2. Schlesinger, M.E. & Mitchell, J.F.B. in *The Potential Climatic Effects of Increasing Carbon Dioxide* (eds MacCracken, M.C. & Luther, F.M.) 81–147 (U.S. Department of Energy, DOE/ER-0237, 1985).
3. Schlesinger, M.E. & Mitchell, J.F.B. *Rev. Geophys.* **25**, 760–798 (1987).
4. Somerville, R.C.J. & Remer, L.A. *J. geophys. Res.* **89**, 9668–9672 (1984).
5. Betts, A.K. & Harshvardan, J. *geophys. Res.* **92**, 8483–8485 (1987).

Contribution of organic acids to lake acidity

SIR—We would like to reply to the correspondence by Krug¹, who raised criticisms of a recent paper² of ours. In that paper, we quantified the contributions of organic acids to lakewater acidity, in the simplest case by comparisons of lakes with similar concentrations of base cations and sulphate but differing in total organic carbon (TOC). The contribution of dissociated organic acids is $4\text{--}5\ \mu\text{eq mg}^{-1}\ \text{TOC l}^{-1}$, a value shown to be independent of Ca^{2+} , SO_4^{2-} and pH when $\text{pH}\sim 5\text{--}7$; the value decreases to $2.5\text{--}3.5\ \mu\text{eq mg}^{-1}\ \text{TOC l}^{-1}$ in lakes with $\text{pH}<4.8$ (ref. 3).

We measured contributions of organic acids to strong and weak acidity. The average pK of organic acids is $3.5\text{--}5.0$ (ref. 4) and there were approximately equal quantities of strong and weak organic acids. The range of pK s indicates that there is no relationship between pH and μeq of charge per mg TOC for $\text{pH}>5$. The organic acids cannot be entirely weak acids as claimed by Krug¹, but act as pH buffers as sulphate concentration increases, particularly if H_2SO_4 is added to water that has a pH near the pK of abundant organic acids.

Krug cites palaeolimnological data for Lake Hovvatn in southern Norway to support his hypothesis that strong mineral acid replaces organic acid on an equivalent basis, but he fails to address the magnitude of the total changes in lakewater chemistry. Whether or not pre-acidification TOC was 2–3 times higher than now is unimportant, but the absolute magnitude of changes in organic anion, base cation, H^+ and aluminium concentrations are significant. A change in pH from 5.2–5.4 to 4.4 represents a $34\text{--}\mu\text{eq l}^{-1}$ increase in H^+ , and this has been accompanied by an increase in Al of $\sim 20\ \mu\text{eq l}^{-1}$. Krug has therefore underestimated both the Al concentration (which influences fish mortality) and the total change in acidity. Studies of diatom reconstructions of lakewater pH indicate that pH changes are associated with atmospheric deposition^{5,6}.

Krug misquotes the value of the factor F , the normalized estimated anthropogenic SO_4^{2-} input (ref. 7). Henriksen⁷ did not propose 0.4 as a regional value, but rather demonstrated that this was the maximum possible value at current acidity levels for concentrations $\text{Ca}^*+\text{Mg}^*=0$; the value $F=0.2$ was proposed as being realistic, in the absence of further data. Clearly, F is a function of base cation concentration, and must approach zero in the most dilute lakes. Neither the organic buffering hypothesis nor the F factor discount the validity of our regression analysis or those of others⁸.

The units of conductivity in our paper were mS m^{-1} rather than $\mu\text{S cm}^{-1}$. The data are of high quality and the ionic

contributions balance.

The empirical data on lakes and streams indicate a correlation between the location of acidic clearwater lakes and the spatial and temporal occurrence of elevated aluminium levels and acid deposition. Krug cannot explain why such lakes are found only in areas of Norway that receive acid rain but not in areas with lower deposition of sulphate. Elevated sulphate levels also lower the pH of humic lakes in areas receiving acid deposition below that from organic acidity.

DAVID F. BRAKKE

*Institute for Watershed Studies,
Western Washington University,
Bellingham, Washington 98225, USA*

ARNE HENRIKSEN

*Norwegian Institute for Water Research,
PO Box 33, Blindern,
N-0313 Oslo, Norway*

STEPHEN A. NORTON

*Department of Geological Sciences,
University of Maine,
Orono, Maine 04469, USA*

1. Krug, E.C. *Nature* **334**, 571 (1988).
2. Brakke, D.F., Henriksen, A. & Norton, S.A. *Nature* **329**, 432–434 (1987).
3. Henriksen, A., Brakke, D.F. & Norton, S.A. *Envir. Sci. Technol.* (in the press).
4. Perdue, E.M., Reuter, J.H. & Parrish, R.S. *Geochim. cosmochim. Acta* **48**, 1257–1263 (1984).
5. Davis, R.B. *Quat. Sci. Rev.* **6**, 147–163 (1987).
6. Flower, R.J., Battarbee, R.W. & Appleby, P.G. *J. Ecol.* **75**, 797–824 (1987).
7. Henriksen, A. *Verh. int. Verein. Limnol.* **22**, 692–698 (1984); *Acid Rain Res. Rep.* no. 1 (NIVA, Oslo, 1982).
8. Gorham, E. *et al.* *Nature* **324**, 451–453 (1986).

Two chimps for HIV

SIR—The 'two chimpanzees' approach endorsed in HIV research by Prince *et al.* (*Nature* **333**, 513; 1988) is challenged by Holland (*Nature* **334**, 478; 1988) on statistical grounds. But it is the only accessible approach, and provided one knows its limitations it is reasonable and indispensable. The numbers calculated by Holland are so high that they would have precluded the development of hepatitis B vaccines. Yet correct inferences were drawn continuously from very small numbers of chimpanzees during the development and production of hepatitis B vaccines.

With HIV, the 'two chimpanzees' approach has shown that several candidate antigens fail to protect against experimental HIV infection, and it has been correctly concluded that one cannot expect significant protection of man against HIV with these formulations. Any others that are developed should be shown to protect at least two chimpanzees before being studied in man. HIV-infected chimpanzees need stricter and lifelong infectious containment than do hepatitis-B-infected chimpanzees — an additional reason for their parsimonious use.

JAN DESMYTER

*Rega Institute for Medical Research,
B-3000 Leuven, Belgium*

Radiation hazards in space put in perspective

SIR—Now that timetables are under consideration for further visits to the Moon and for an exploratory mission to Mars, the Commentary by Letaw *et al.*¹ on the radiation hazards of space travel should be put into perspective.

The radiation environments in space are certainly less benign than on Earth, but Letaw *et al.* paint an unnecessarily gloomy picture for missions beyond the magnetosphere. They have raised important questions about the shielding and secondary radiations generated by the interaction of the primary particles and the shielding material, as well as the dangers of solar particle events. We are concerned that they may have given the impression that the dangers from radiation, especially from heavy ions and the infrequent very large solar particle events, are not only frighteningly large but known with confidence.

As Letaw *et al.*¹ aver, little is known as yet about the effects of heavy ions on humans, so information must be gleaned from experiments with animal models. These authors state that because of the comparative deposition of energy "a rare cosmic-iron nucleus is 13,500 times as dangerous as a cosmic ray proton". Yet, the available assessments of the comparative biological dangers of radiations of different qualities have not been based on fluence but on dose. The relationships between comparative energy deposition and comparative danger are complex. Relative biological effectiveness (RBE), based on dose, for a number of endpoints, such as life shortening, cataract and cancer, is in the range of 1 to about 30 (refs 2–8; E. Blakely, personal communication).

Furthermore, it should be noted that the fluence of iron particles is low. Not more than about one per cent of galactic cosmic rays are heavy ions. In low Earth orbits the contribution of galactic cosmic rays to the total dose equivalent is small. For example, in the orbit proposed for the Space Station the galactic cosmic ray dose equivalent is estimated to be about 7 mSv (0.7 rem) for a ninety-day mission, or a little more than twice the average annual effective dose equivalent from various sources of natural background radiation on Earth⁹. This is not to imply that the radiation levels are of no concern. They are of concern but they are manageable.

Apart from the effects of galactic cosmic rays and the secondary radiations in missions beyond the magnetosphere, the authors state that "rare solar particle events could deliver a lethal dose of up to 1,000 rem in one day". Such high dose equivalents could be incurred but, to our knowledge, only in the skin. Major contributors to such high surface doses are low-energy protons that do not penetrate into

the life-limiting tissues such as gut and bone marrow. Other estimates of the marrow dose equivalents for the anomalously large solar particle events in 1972, with an assumed shielding of 2 g cm^{-2} , are about 1 Sv (100 rem). One sievert is a high dose equivalent but not lethal. Furthermore, there is little reason to believe that any astronaut would be exposed for one day to a major solar particle event without considerable shielding.

The authors point out that their estimate of the dose equivalent for a very large solar particle event is higher than other estimates^{10,11} and attribute the difference in estimates to the inclusion by them of such secondary particles as neutrons. It appears that Letaw *et al.* have used a quality factor of 20 for neutrons; this is unjustified in considering acute damage to the gut and bone marrow. Quality factors are based on RBE for stochastic late effects, such as cancer, and not acute effects; for neutron-induced acute effects involving cell killing, the RBE is 2–4, not 20.

Letaw *et al.*¹ state that the maximum galactic cosmic-ray dose coupled with other radiation exposures exceeds the annual dose limit of 0.5 Sv (50 rem) recommended by the National Council on Radiation Protection and Measurements (NCRP)¹². But the full recommendation is that 50 rem should not be exceeded in any one year and that the career limits for astronauts should be 200 to 400 rem depending on age and sex. Furthermore, the new guidelines are intended for missions in low-Earth orbits and relatively brief missions in geostationary orbits. For quite some time, missions to the Moon and Mars will be considered exploratory and, from the point of view of radiation protection, in a different category from the Space Station. More information is required about the late effects of radiations to be encountered in deep space.

The NCRP recommendations for career limits are based on the belief that cancer is the principal risk, and are made with an awareness of the risk of space travel itself. The aim is to keep lifetime cancer risks comparable to lifetime occupational risks of activities such as agriculture and construction. Those occupations are neither the safest nor the most dangerous and carry a lifetime risk of about three per cent. NCRP's recommendations are based on the same lifetime risk¹². The 0.5 Sv limit for any one year is aimed at preventing the induction of non-cancer effects.

R. J. M. FRY

Biology Division,
Oak Ridge National Laboratory,
Oak Ridge, Tennessee 37831-8077, USA

J. T. LETT

Department of Radiology,
Colorado State University,
Fort Collins, Colorado 80523, USA

1. Letaw, J.R. *et al.* *Nature* **330**, 709–710 (1987).

2. Kraft, G. *et al.* *Adv. Sp. Res.* **6**, 127–136 (1987).
3. Alpen, E.L. & Powers-Risius, P. *Radiat. Res.* **88**, 132 (1981).
4. Yang, T.C. *et al.* *Radiat. Res.* **104S**, 177–187 (1985).
5. Lett, J.T. *et al.* *Adv. Sp. Res.* **6**, 295–303 (1987).
6. Worgul, B.V. *Adv. Sp. Res.* **6**, 285–293 (1987).
7. Fry, R.J.M. *et al.* *Radiat. Res.* **104S**, 188–195 (1985).
8. Ainsworth, E.J. *Adv. Sp. Res.* **6**, 153–165 (1987).
9. *Ionizing Radiation Exposure of the Population of the United States* Rep. 93 (NCRP, 1987).
10. Curtis, S.B. *et al.* *Adv. Sp. Res.* **6**, 269–274 (1987).
11. Atwell, W. *et al.* *Terrestrial Space Radiation and Its Biological Effects* (eds McCormack, P.D., Swenberg, C.E. & Bucker, H.) (Plenum Press, New York, in the press).
12. *Guidance on Radiation Received in Space Activities* (NCRP, in the press).

LETAW *et al.* reply—We thank Fry and Lett for their clarification of the NCRP deliberations on astronaut radiation dose limits. But their comments have not identified any serious errors in our analysis and do not alter our recommendations.

The factor of 13,500 is derived according to established radiation protection methods. The dose attributed to monoenergetic ions is the fluence multiplied by the linear energy transfer (LET) or stopping power¹. Since the LET of relativistic ions is proportional to the square of the charge², the dose from iron ions is $26^2=676$ times greater than the dose from an equal fluence of protons. The dose equivalent of a monoenergetic fluence of ions is equal to the dose times the quality factor¹. International convention³ sets the quality factor of relativistic iron at 20 and relativistic protons at 1. As documented by Fry and Lett, this is in reasonable agreement with the observed relative biological effectiveness for iron nuclei. The dose equivalent of relativistic iron is therefore $676 \times 20=13,520$ times greater than the dose equivalent of relativistic protons. For endpoints where the dose response is assumed to be linear and stochastic, such as leukaemia, risk is proportional to the dose equivalent⁴. Thus, relativistic iron ions carry a risk of cancer 13,520 times greater than relativistic protons.

The ratio of iron nuclei to protons in cosmic radiation, about 2.5×10^{-4} , was included in our analysis. Iron makes up 25 per cent of the dose equivalent from galactic cosmic radiation with thin shielding, while only 6 per cent of the dose equivalent is due to protons.

Our Commentary addressed space missions of long duration outside the magnetosphere. However, we specifically stated that on the Space Station the radiation risk arises from trapped protons, not galactic cosmic radiation. The annual dose equivalent from trapped protons at an orbital altitude of 500 km is comparable to the NCRP annual limit and is of concern to Space Station planners.

The two 'worst-case' scenarios for solar energetic particle events — the 'composite' worst-case event⁵ and ten times the August 1972 event — result in whole-body doses exceeding the lethal (LD_{50}) value of 350 rads⁶. The August 1972 flare would have deposited a whole-body dose exceeding the NCRP monthly limit of 25 rem in one day. For these reasons we specifically

recommended a storm shelter of at least 9 cm aluminium (or equivalent) for any mission outside the magnetosphere. Fry and Lett state that "there is little reason to believe that any astronaut would be exposed for one day to a major solar particle event without considerable shielding", and therefore largely concur with our recommendation.

We agree with Fry and Lett that conventional quality factors are too conservative for estimating the risk of acute radiation effects, and in our Commentary we specifically noted a factor of two uncertainty in the calculations. Our recommendations took this limitation into account.

We are enthusiastic supporters of manned space flight. However, we are confident that harm to the astronauts and programme setbacks due to radiation hazards are best avoided by careful assessment of the risks before missions are defined and spacecraft designed. In this connection, we initiated spacecraft-shielding studies reported in a paper presented at a recent meeting. We found that hydrogen and other light shielding materials are up to 4.3 times more effective than aluminium at protecting astronauts from galactic cosmic radiation.

JOHN R. LETAW

Severn Communications Corporation,
PO Box 544, Severna Park,
Maryland 21146, USA

REIN SILBERBERG

C. G. TSAO

E. O. Hulbert Center for Space Research,
Naval Research Laboratory,
Washington, DC 20375, USA

1. *Radiation Quantities and Units* (ICRU Rep. 33, International Commission on Radiation Units and Measurements, Washington, DC, 1980).
2. Ahlen, S.P. *Rev. Mod. Phys.* **52**, 121–173 (1980).
3. *Recommendations of the ICRP* (ICRP Rep. 26, Pergamon, Oxford, 1977).
4. Eisenbud, M. *Environmental Radioactivity* (Academic, Orlando, 1987).
5. Adams, J.H. & Gelman, A. *IEEE Trans. Nucl. Sci.* **NS-31**, 1212–1216 (1984).
6. *Radiobiological Factors in Manned Space Flight* (ed. Langham W.H.) (Publ. 1487, National Academy of Sciences, Washington, DC, 1967).
7. Letaw, J.R., Silberberg, R. & Tsao, C.H. *Adv. Space Res.* (in the press).

Malapropism?

SIR—While commending the original observations, one questions the new use of the term "piracy" in describing surrogate paternity by adulterous wrasses that cuckold smaller, conspecific nesting fishes (E. P. Van den Berghe, *Nature* **334**, 697–698; 1988).

Piracy by frigate birds refers correctly to the stealing of booty. Could not van den Berghe have chosen a more appropriate term for the wholly different phenomenon he describes in wrasses?

RALPH A. LEWIN

Scripps Institution of Oceanography,
University of California,
San Diego, La Jolla,
California 92093, USA

The armchair balloonist

Gordon L. Herries Davies

Theories of the Earth and Universe: A History of Dogma in the Earth Sciences.
By S. Warren Carey. *Stanford University Press: 1988. Pp.413. \$45.*

POLYGENOUS is the first adjective to apply to this book. Professor Carey — a doyen among Australian geologists — has given us a work that is part autobiography, part history of science, part treatise upon global tectonics and part speculative cosmology. It is indeed a rare amalgam which Carey himself describes as “this armchair philosophy of an aging scientist”. But despite its divers themes, the book does have an overall cohesion. Earth scientists will find it a thought-provoking work even though it will raise many a hackle.

Carey is convinced that the modern theory of plate tectonics is a false dogma. He accepts the theory in so far as it invokes seafloor spreading, but he rejects the notion that a balance is maintained between the creation of new crust along the mid-ocean ridges and the destruction of old crust along zones of subduction. For Carey, subduction is a myth. In his eyes the Earth's surface is not the site of a steady-state economy. There is no need, he claims, for the accruing income at the mid-ocean ridges to be balanced by the expenditure of subduction, because the Earth, like a great balloon, is steadily expanding at an exponential rate. For most geologists, Carey's views are neither welcome nor acceptable. That eminent professor who at the close of one of Carey's expositions rose and exclaimed “Bull-shit!” may have wanted manners but he expressed the consensus of the scientific community. Carey nevertheless remains unabashed.

The book opens with obeisance to Clio. Carey seeks to show that his own position is by no means unique. He claims that throughout the history of the Earth sciences prevailing dogma has repeatedly hindered the acceptance of important new ideas. He offers as exemplars of his thesis such familiar cases as the debate over the nature of fossils, the controversy between the neptunists and the plutonists, and the battle that raged around the glacial brainchild of Louis Agassiz. This is the least satisfactory part of the book. My complaints are twofold. First, there are a large number of small factual errors, and secondly — and far more importantly — I found Carey's whiggish interpretation of history to be quite unacceptable. He presents the history of his science in terms of ‘baddies’ and ‘goodies’. He sees his baddies as the blinkered obstructionists who clung to outmoded dogma and who

thus hindered the forward march of science. His goodies were those who, in his eyes, struggled valiantly to establish what we now hold to be scientific truth.

This simplistic interpretation will not do. Carey's baddies were never the myopic dunderheads that he implies. Rather they were scientists just as able as any of our own day and equally resolute both in their search for truth and in their combating of what they saw as error. Carey throws one of his darts at William Richardson, that wernerian geologist who, early in the last century, refused to see the significance of a site shown to him by some huttonians upon Edinburgh's Salisbury Crags. But please reflect. Richardson was an Irish geologist deeply experienced in the basaltic landscapes of his native land. Why should this demonstration of a small piece of sandstone set within trap be expected to shake a man's long-cherished and carefully reasoned belief in the wernerian theory? Would Carey allow just one geological section to subvert his own 30-year-old belief in an expanding Earth? The relationship between what the geologist sees and what the geologist believes is far more complex than Carey's historical review would suggest.

In the second part of the book, Carey outlines the history of the theories of continental drift and plate tectonics. This is now familiar ground and my only comment is to regret that Carey here perpetu-

ates the notion that Francis Bacon was the first to point out the ‘jig-saw’ fit of the opposing shores of the Atlantic. Bacon did no such thing. He merely commented that the Old and New Worlds have similar shapes in that they both taper southwards, and in that the outlines of Africa and South America are similar if we compare their equivalent coasts, viz. the ‘horn’ of East Africa with the ‘shoulder’ of Brazil and the Gulf of Guinea with the Peru-Chile bight.

The burden of Carey's thesis is presented in Parts 3, 4 and 5 of the book. There he forcefully presents his case for belief in an expanding Earth, dealing with certain concomitant matters (especially with what he terms “vertical orogenesis”) and countering some of the objections which can be raised against his theory. The final part is devoted to cosmological speculations in such areas as Big Bang theory (which Carey rejects), antimatter, black holes and gravity waves. Here, as ever, Carey's writing is both clear and stimulating, although among cosmologists I suspect he will find more opponents than friends.

In publishing this work Carey breaks one of the unwritten rules of the world of science. He claims in his preface that he is addressing “the educated layman” and I am sure that he will be widely read within that sector of society. Here he is making an appeal to the general public over the heads of his scientific peers. But then the same was true of the theory of continental drift in 1915, when Alfred Wegener published his book championing that idea. Today Wegener's *Die Entstehung der Kontinente und Ozeane* is a revered classic of the Earth sciences. Will Carey's book one day be held in similar esteem? □

Gordon L. Herries Davies is a Fellow of Trinity College, Dublin 2, Ireland.



Heterodox precedent — Alfred Wegener (1880–1930), whose ideas about continental drift, published in the early part of this century, presaged the acceptance of the concept 50 years later as new data were brought to bear on the problem. The picture is taken from the second edition of Earth, Time, and Life: An Introduction to Physical and Historical Geology by Charles W. Barnes, just published by Wiley, price £35.55, \$55.10.

ADVERTISEMENT

The ANS Journals

Your number one source for current information about
NUCLEAR ENERGY

NUCLEAR NEWS

NUCLEAR SCIENCE & ENGINEERING

NUCLEAR TECHNOLOGY

TRANSACTIONS

REMOTE SYSTEMS TECHNOLOGY PROCEEDINGS

FUSION TECHNOLOGY

For free catalog contact:

AMERICAN NUCLEAR SOCIETY
555 N. Kensington Avenue
La Grange Park, Illinois 60525 USA
312-352-6611 Telex: 4972673

Metastatic Spread of Cancer

G.V. Sherbet

This book outlines a new approach to the analysis of the mechanisms of metastasis by comparing the glioma, an intracranial tumour which ordinarily does not metastasize, to other metastasizing tumour models. This has produced a coherent picture of the operation of oncogenes and growth factors in the initiation, development and progression of neoplasms. This is an essential book for those researching metastasis and will be of interest to all research oncologists.

1987 £45.00 275pp
234x156mm 0 333 443 624

Order this book through your bookseller or send your order to:
Richard Gedye, Macmillan Press Ltd,
Houndmills, Basingstoke RG21 2XS, UK.

To the heart of an ancient empire

Arthur A. Demarest

The Great Temple of the Aztecs: Treasures of Tenochtitlan. By Eduardo Matos Moctezuma. Translated by Doris Heyden. *Thames & Hudson: 1988. Pp.192. £24. In the US available from W.W. Norton, \$29.95.*

IN THE early sixteenth century, Hernan Cortez and his conquistadors destroyed the ancient city of Tenochtitlan, the capital of the sprawling Aztec empire. The Spanish destroyed or buried most of the city's splendid temples and palaces. Over the ruins of Tenochtitlan they raised their own colonial capital, Mexico City, from which to govern the newly subjugated regions and to provide the base for their further conquests and their eradication of Aztec idolatry and cults of human sacrifice.

Four and a half centuries later, in the 1970s, tunnel construction for Mexico City's subway system initiated one of the most exciting episodes in modern archaeology, the excavation of the Great Temple of the Aztecs. The fortuitous discovery of several monumental sculptures led to the location of the base of this legendary structure, the religious centre of fifteenth- and early-sixteenth-century Mesoamerica. Mexican archaeologists then began a five-year project to uncover the base of the Great Temple and explore the history of its construction.

In *The Great Temple of the Aztecs*, the leader of that project, Eduardo Matos Moctezuma, describes the excavations and the discoveries in the legendary Aztec 'Templo Mayor'. In the book, the latest in the series *New Aspects of Antiquity*, he tries to place the Great Temple and its treasures in the context of the history and nature of the Aztec empire. The result is a fascinating initial view of these important archaeological finds for both the scholar and the general reader.

Matos Moctezuma begins with an overview of the prehistory of central Mexico and the history of archaeological research in Mexico City. The general reader may find his account of the prehistory too brief and schematic, and that of the archaeological researches unnecessarily detailed. But the author finds his stride when he turns to his own area of expertise, the history of the Aztecs and the nature of their fourteenth- and fifteenth-century civilization. These chapters provide the background for the core of the book — the description of the Great Temple and its treasures.

Chapters dealing with the excavations and artefacts are simply written, in clear

undramatic style, but the nature of the finds requires no literary embellishment. Excellent and numerous black-and-white photographs, colour plates and drawings show the temple architecture, monumental sculptures, grisly human sacrifices, and the hundreds of beautiful art objects recovered from offerings buried within the structure. The wealth and power of the Aztec empire are displayed in both the richness and the surprising geographical diversity of the stone, bone, metal and ceramic objects discovered. Most of the treasures were imported from distant regions such as Pacific coastal Guerrero, Veracruz and highland Oaxaca. Particularly common are idols and braziers of the principal deities of the temple — the rain god, Tlaloc, and Huitzilopochtli, the Aztec patron deity of war, human sacrifice and the Sun.

In other, speculative chapters the author shows that details of the temple's architecture and offerings represent the myths of Huitzilopochtli and the role of the temple as the cosmic centre of the Aztec universe. Scholars may take issue with some of the interpretations here. The presentations of Aztec history seem, at times, flatly to accept the official 'state' versions of the Aztecs' rise to imperial power. The identification and emphasis on the fire deity, Xiuhtecutli, in temple offerings, and interpretations of the myths of the underworld are speculative — as the author admits.

Still, these controversies merely add to the sense of immediacy as scholars struggle with the overwhelming corpus of art and imagery found during this extraordinary excavation. □

Arthur A. Demarest is Centennial Professor of Anthropology at Vanderbilt University, PO Box 6050-B, Nashville, Tennessee 37235, USA.

New in paperback

- *The Making of the Atomic Bomb* by Richard Rhodes. Publisher is Penguin, price is £7.99. For review see *Nature* 328, 301 (1987).
- *Missing Links: the Hunt for Earliest Man (2nd edn)* by John Reader. Publisher is Pelican, price is £4.99, \$7.95. For review see *Nature* 291, 599 (1981).
- *Time's Arrow, Time's Cycle: Myth and Metaphor in the Discovery of Geological Time* by Stephen Jay Gould. Publisher is Pelican in the UK, price £3.99; Harvard University Press in the US, price \$8.95. For review see *Nature* 326, 898 (1987).
- *The Wonderful Mistake: Notes of a Biology Watcher* (incorporating *The Lives of a Cell* and *The Medusa and the Snail*) by Lewis Thomas. Publisher is Oxford University Press, price is £6.95. For reviews see *Nature* 258, 24 (1975) and 282, 140 (1979).

New in Britain

Published today by Unwin Hyman —

- *Superconductors: The Breakthrough* by Robert M Hazen. Price is £12.95. For review see *Nature* 334, 479 (1988).
- *The Statue Within* by François Jacob. Price is £12.95. For review see *Nature* 328, 389 (1987).

New order for the apiary

John B. Free

Biogeography and Taxonomy of Honeybees. By Friedrich Ruttner. *Springer-Verlag*:1988. Pp.284. DM 158, £53, \$87.50.

PROFESSOR Friedrich Ruttner is the foremost authority on the taxonomy and distribution of honeybees. He has now presented us with a masterly account which incorporates the results of many years of his own research and study into the evolution, taxonomy, ecological adaptation, geographical variability and comparative behaviour of the four honeybee species in the sub-family Apinae. This is the first one-volume overview of the subject.

The classification of honeybees, and especially of their sub-species, has been a subject of continuous discussion and argument. Indeed, Ruttner reminds us that Linnaeus himself first described the Western honeybee as *Apis mellifera* in 1758 and three years later changed the name to *A. mellifica*. Traditional methods of systematics, which proved to be confusing and inadequate, have now been largely replaced by effective methods of morphometric analysis of honeybee populations.

Ruttner's own results and conclusions are based on the more than 1,400 population samples in the morphometric bee databank at the Institut für Bienenkunde, Oberursel, of which he was formerly director. Many of these samples, particularly those collected by Brother Adam of Buckfast Abbey, are probably now irreplaceable because of man's movement of honeybee colonies from one region to another. To discriminate between samples up to 40 characters may be used, and lists of these characters, the methods of their selection, measurement and statistical analysis are provided. Clear discrimination between the bees of two races is important for practical bee-keeping and bee-breeding programmes. Fortunately, a few characters only may suffice for particular purposes, such as distinguishing between Africanized and European colonies in South and Central America.

Compared to *A. mellifera*, the three Eastern honeybee species (*A. cerana*, *A. dorsata* and *A. florea*) have been little studied; for example, the only information Ruttner was able to provide on such an important subject as the division of labour in *A. cerana*, the sole Eastern species kept in hives, was from unpublished observations made by a student in a 'flight room' in Germany! Nonetheless, interesting advances have been made in

the past decade or so, and the comprehensive reviews of the existing literature on these species are especially welcome and serve to illustrate the need for more research in order to take advantage of their full economic potential as pollinators and honey-producers.

For many readers the prime value of the book will be the clear presentation of the classification, characterization and distribution of the 24 races of *A. mellifera*. These range from races that can be maintained throughout the year in the Arctic circle to those kept in the extremely hot and arid conditions of Arabia. Such adaptation to the environment must not be jeopardized by importing honeybee races from other areas. The common European black bee, for instance, is now heavily hybridized and has vanished completely

from many areas in central Europe; Ruttner found that the imported black English bee living in the trees and cliffs of Tasmania is more 'typical' of the race than the present descendants of their common ancestor in Britain.

We can hope that definition of the taxonomic unit to which a colony belongs will eventually allow accurate prediction of the colony's behaviour. In the past many research workers have given too little attention to the particular race of honeybee they were studying. Until recently the classification was so confused that there was some excuse; that is no longer so. □

John B. Free is Honorary Professorial Fellow in the School of Pure and Applied Biology, The University of Wales College of Cardiff, Cardiff CF1 1XL, UK.

Mountains of facts

Keith Bell

McGraw-Hill Encyclopedia of the Geological Sciences, 2nd edn. Editor-in-chief Sybil P. Parker. *McGraw-Hill*:1988. Pp.722. \$85, £65.

AN ENCYCLOPAEDIA should be easy to use, informative and up to date. On these counts, the new edition of the *McGraw-Hill Encyclopedia of the Geological Sciences* is only partly successful.

The information is well set out in a two-column format and the illustrations, both line drawings and photographs, are of a uniformly high quality. But although the overall presentation is pleasing, the organization of the material is less satisfactory. Finding a complete run-down of a particular topic is not an easy matter; time and persistence are needed. Related subject matter tends to be scattered throughout the volume so that a great deal of rummaging around and background knowledge is needed before the information can be successfully extracted. Aspects of geochronology are dealt with under at least six different headings, and although most of the thumb-nail sketches such as "Earth, age of" and "Rock age determination" are succinct, clearly written and enjoyable reading, there is considerable overlap and repetition from one section to another. Although many entries also lead the reader to other, related topics, and the index is generally useful, there are some surprising omissions — for example, I could find neither fossils nor palaeontology (or rather paleontology) included either among the main headings or in the index.

Most of the main aspects of geology are covered, including evolution of continents, marine geology, metamorphism and petroleum geology, to name but a

few. A quick search, however, for more specific topics, failed to reveal any information about komatiites or pyrochlore. It took some time to track down the trilobites, tucked away in a few lines under the entry for Cambrian.

There is, too, a dated air about many parts of the *Encyclopedia* — the most recent reference for palaeomagnetism is 1979, and the references for several other topics could have been more up to date. There is no excuse for including a map showing the arrangement of mountains and continents of the world (without a legend!) dated 1935.

It's easy to criticize a volume containing material of such an interdisciplinary and wide-ranging nature, and the book does have its uses. It will serve as a general reference work or for browsing; it contains a great deal of useful information and the entries are consistently well written in a clear, concise style. Some sections, such as those on feldspar and geological thermometry, are excellent. But for the general reader interested in tracking down a specific subject, or for the researcher wanting the latest information on a topic, the book is decidedly deficient.

Most of the material was previously published in the sixth (1987) *McGraw-Hill Encyclopedia of Science and Technology* and reorganized into the present format, which might explain some of the weaknesses. If the editors are to make the next edition an essential acquisition for Earth scientists, careful pruning and pulling together of related subject matter will be needed, as well as better judgement over establishing the relative importance of the various topics — in the present volume, for example, continental drift merits half a page, amber a page and a half. □

Keith Bell is Director of the Ottawa-Carleton Geoscience Centre, Department of Earth Sciences, Carleton University, Ottawa K1S 5B6, Canada.

Programs for quicker relationships

Norman I. Platnick

Hennig86, 1.5. By J.S. Farris, 41 Admiral Street, Port Jefferson Station, New York 11776. *IBM-PC and compatibles*. \$50.

PAUP, 3.0. By D.L. Swofford, 607 East Peabody Drive, Champaign, Illinois 61820. *Apple Macintosh, and IBM-PC and compatibles*. \$50. *Upgrades* \$25.

HYPOTHESES about the phylogenetic relationships of organisms are central to many aspects of evolutionary biology, from systematics and biogeography to studies of speciation and evolutionary rates. Most systematists today group species into higher taxa (genera, families and so on) solely by their degree of phylogenetic relationship. This approach is commonly called cladistics, because its proponents include in classifications only clades — groups of species that are all more closely related to each other than any of them are to species excluded from the group.

The evidence for clades consists of characters that are unique to the groups thus constituted, such as mammary glands (unique to mammals) or spinnerets (unique to spiders). In practice, a matrix of taxa and their characters is constructed, and that information is used to infer a branching diagram (cladogram) specifying all the clades supported by the data. In some cases, usually involving relatively small numbers of taxa and characters that are in complete (or nearly complete) agreement, it is possible to construct a cladogram by hand. Indeed, the cladograms produced by the founder of cladistics, the late German entomologist Willi Hennig, were all inferred entirely by hand. Those inferences involve a parsimony procedure: joint possession of a character is taken to be evidence of common ancestry unless other characters disagree with that grouping.

Larger data sets

For larger data sets, or those in which several characters disagree, discovery of the correct cladogram(s) by hand is impractical. For just nine taxa, for example, there are over two million possible cladograms, even counting only those arrangements that are completely specified (that is, are dichotomous, having but two branches arising from each node). Moreover, the problem of finding the most parsimonious (shortest) cladogram for a given data set has been shown to be NP-complete, meaning that an efficient algorithm which will always generate the correct solution is unlikely to exist. Given these difficulties, it is hardly surprising that cladists routinely depend on computer programs to analyse their data.

The two new programs reviewed here — the 1.5 edition of *Hennig86* and the 3.0 edition of *PAUP* (Phylogenetic Analysis Using Parsimony) — both run on micro-

computers and provide an enormous increase in the accuracy and efficiency with which solutions can be obtained. Each of them implements both exact and approximate algorithms.

The exact algorithms use a branch-and-bound strategy to eliminate suites of cladograms that cannot possibly be among the shortest arrangements. Branch-and-bound implementations in previous microcomputer programs* have been either too slow for problems involving more than a dozen taxa, or have produced output requiring hours of hand analysis. For example, the *PENNY* program in J. Felsenstein's *PHYLIIP* package, and an earlier version of *PAUP*, each produce 54 cladograms for a data set involving 12 taxa and 42 characters. But not all of those cladograms are different; although they are treated as completely dichotomous, some of their branches are unsupported by characters (are of zero length), and when those branches are collapsed into polychotomies, identical cladograms often result. Several hours of work are required to determine that of the 54 cladograms, only 19 are distinct. The new programs produce just those 19 distinct cladograms, with no redundancy. The older programs could store and output only 100 cladograms, even if many more were discovered, and because of redundancy those 100 were often an insignificant sample of the actual range of equally parsimonious solutions. The new programs can store many more cladograms (2,500 or more, depending on the number of taxa), all of which are really different.

The efficiency with which these exact algorithms work is data-dependent. For data sets with little disagreement among the characters (that is, with an internal consistency near 1.0), astonishingly quick results are possible; both *PAUP* and *Hennig86* obtained exact answers for a 21-taxon data set with a consistency index of 0.82 in less than 15 seconds. *Hennig86* solved a noisy 20-taxon data set (consistency index 0.38) in just four hours — almost three times as fast as the best previous program, and with complete solution of 90 different cladograms, rather than partial solution of 100 with duplicates.

The approximate algorithms make initial estimates of the best cladograms and then systematically rearrange their branches in search of improved arrangements. Unlike the branch-and-bound approaches, these algorithms are not

guaranteed to produce optimal results, but in practice they do extremely well, and run very quickly; with large data sets they are the only feasible approach. Each program offers many combinations of options for repeating the analyses with different initial estimates of the correct cladogram, providing diligent users with exceptionally good chances of finding optimal solutions. For example, out of 15 data sets that are too large for exact solutions to have been found, *Hennig86* and the new *PAUP* each found, in five cases, results better than had been obtained from any previously available mainframe or microcomputer program.

Greater weight

Both programs allow users to give greater weight to some characters than to others, and to treat multi-state characters as additive (such that changing from state A to C requires an intermediate change to state B) or non-additive (such that any state can transform directly into any other). I ran a total of 86 analyses on *Hennig86*, involving all combinations of weighted/unweighted and additive/non-additive characters, in which the results of four different approximate options were compared to an exact solution. In every case at least one of the approximate options found the shortest cladogram. Moreover, there was no instance in which at least one of the approximate options did not find all of the shortest cladograms.

Both programs are loaded with features that will endear them to users. *PAUP* includes intuitive pull-down menus, and a built-in editor with syntax error pointing; the Macintosh version will interface directly with the forthcoming update of the excellent *MacClade* program, which allows interactive cladogram rearrangement and character-change tracing on the Apple machines. *Hennig86* includes a smooth implementation of successive approximations weighting, allowing users to determine which of the equally parsimonious cladograms are supported by the least noisy characters, and has its own *MacClade*-like facility for interactive tree rearrangement and character tracing.

As for efficiency, *Hennig86* on an 80386-based PC analysed a suite of 25 data sets over three times faster than the older *PHYSYS* program on a CYBER mainframe, even though it was saving ten times as many cladograms and finding better solutions for six data sets. The moral — dump the mainframe, get a fast PC and enjoy using these programs. □

*Earlier programs were discussed by M. Luckow and R.A. Pimentel in *Cladistics* 1, 47–66 (1985); W.L. Fink in *Science* 234, 1135–1139 (1986); and myself in *Cladistics* 3, 121–144 (1987).

Norman I. Platnick is Chairman of the Department of Entomology at the American Museum of Natural History, New York, New York 10024, USA.

The functional logic of cortical connections

S. Zeki & S. Shipp

Department of Anatomy, University College, Gower Street, London WC1E 6BT, UK

Patterns of anatomical connections in the visual cortex form the structural basis for segregating features of the visual image into separate cortical areas and for communication between these areas at all levels to produce a coherent percept. Such multi-stage integration may be a common strategy throughout the cortex for producing complex behaviour.

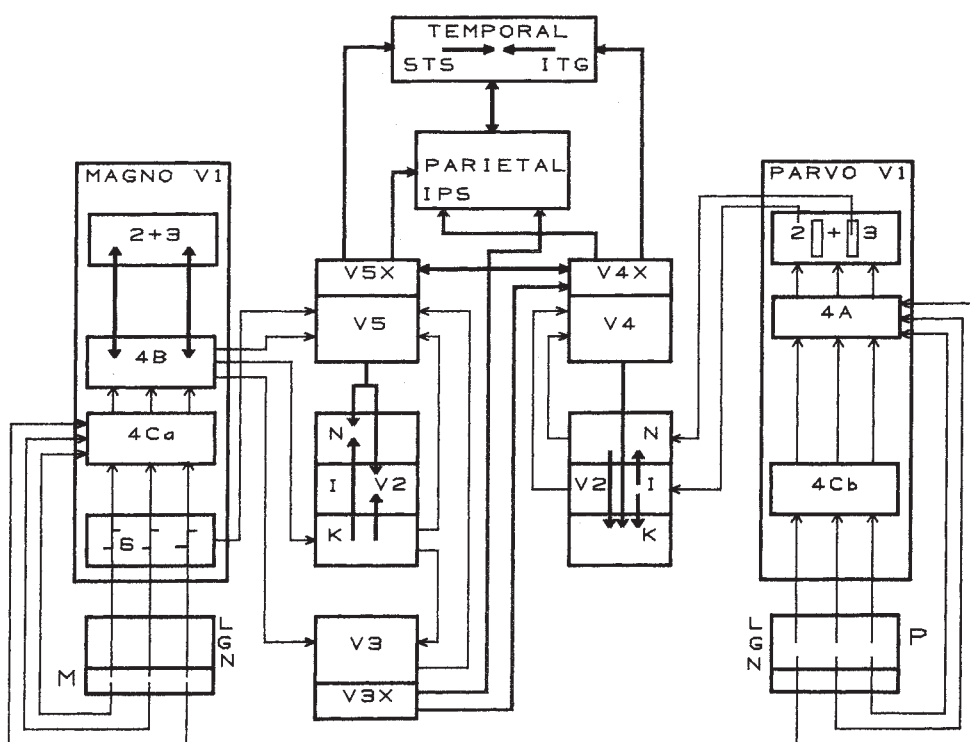
It is an everyday experience that the visual world is perceived as a unified whole and that all the various properties of objects—their distances, shapes, sizes, colours, and directions of motion—are seen together in exact spatio-temporal registration. Yet this effortless coherence within the visual image bears little trace of the internal subdivisions of function in the visual cortex¹ that have become apparent from the past twenty years of research into the primate visual system. Anatomical, physiological and behavioural experiments in the monkey¹⁻⁴ and clinical studies of humans with cerebral lesions^{5,6} have established that different attributes of the visual scene, such as form, colour and motion, are processed in separate, anatomically distinct regions in the visual cortex, each executing its functions with considerable autonomy. Bilateral damage to a particular part of the human cerebral cortex, for instance, can result in a complete loss of the sensation of colour⁵ or of motion⁶, without any other notable deficit. In short, these findings support a theory of functional specialization in the visual cortex¹.

This separation of functions mirrors that found in the cerebral

cortex as a whole, where increasingly large numbers of specialized areas dealing with sensory and motor functions have been discovered. But with every addition a major problem, elegantly phrased by Karl Lashley⁷, has become increasingly apparent: "... how [it is that] the specialized areas of the cerebral cortex interact to produce the integration evident in thought and behavior". In effect the discovery of specialization renders this problem of integration more acute. In the older concept of cerebral organization, that of a single hierarchical chain of areas, integration was envisaged to occur progressively in the transition from one area to the next, each carrying out a higher level of analysis of all the same attributes than its predecessor^{8,9}. But it is now established that there are several serial pathways running in parallel, each of which is functionally specialized; these are actually physically segregated from each other and integration thus demands some form of communication between them.

We survey here the different strategies by which the visual cortex achieves both segregation and integration. They involve both forward and backward connections along a serial pathway,

Fig. 1 Diagrammatic representation of the projections of the P and M pathways to the specialized areas of the striate and prestriate visual cortex (black arrows) and the interconnections between them (coloured arrows). Forward connections are shown in green; like the connections shown in black, they are all reciprocated by a backward projection but only the forward direction is indicated. Connections in red are backward and those in blue are lateral or intrinsic. Interconnections between the specialized pathways and areas exist at many different levels but the early stages of the pathways are shown in greater detail because they have been more extensively studied than the later stages where the definition of cortical areas and their interconnections is not yet complete. The diagram shows the lateral geniculate nucleus (LGN) subdivided into magnocellular (M) and parvocellular (P) layers; V1, subdivided into layers 6, 4Cb, 4Ca, 4B, 4A, 3 & 2, with cytochrome oxidase blobs (small cylinders) in layers 2 & 3 in the P system on the right; V2, subdivided into cytochrome oxidase thick (K), thin (N), and inter (I) stripes; areas V3, V4 and V5 as part of the third, fourth and fifth visual complexes denoted by V3X, V4X and V5X (connections shown to arise from V3X, V4X and V5X include the whole complex and do not exclude the areas V3, V4 and V5 proper); and the higher visual areas in the parietal and temporal lobes, the former including the intraparietal sulcus (IPS), among other subdivisions, and the latter the superior temporal sulcus (STS) and inferotemporal gyrus (ITG), all of which probably consist of several separate areas.



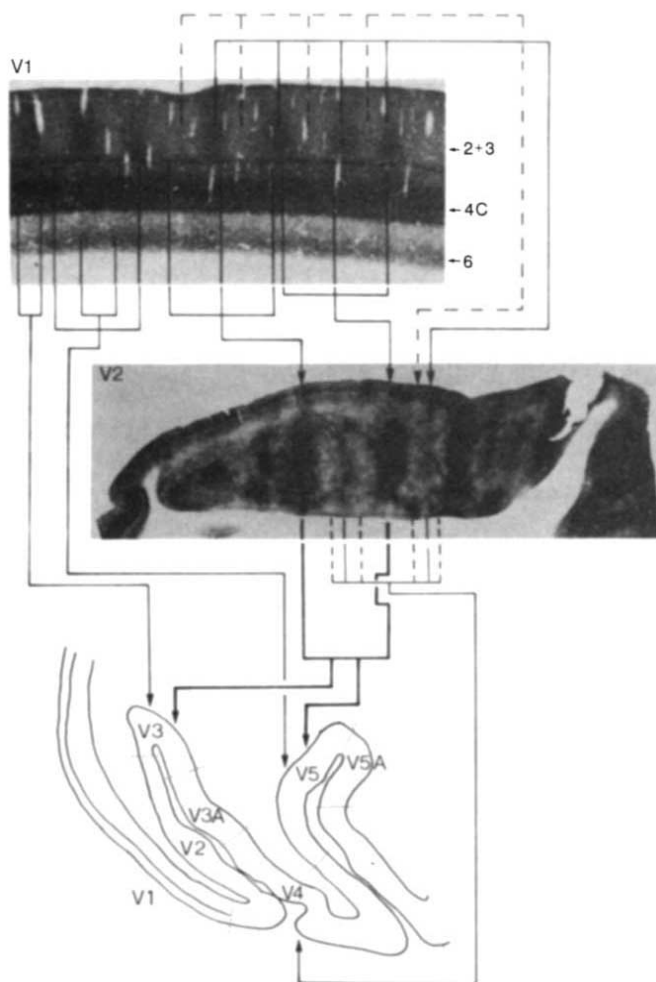


Fig. 2 The separate pathways leading from V1 through V2 to specialized areas V3, V4 and V5 in the prestriate cortex, shown by a stain for cytochrome oxidase activity. Sections through V1 and V2 taken at angles which most clearly reveal their cytochrome oxidase architecture are shown at the top and centre respectively. The M pathway includes the output from layers 4B and 6 directly to V5 and V3 and indirectly to the same areas through the thick dark stripes of V2 (heavy lines). The dark blobs and lighter interblobs of the upper layers of V1 are part of the P pathway. The blobs project through the thin dark stripes of V2 to area V4 (thin lines) and the interblobs project through the paler interstripes of V2 to V4 (dashed lines). A smaller direct projection from the upper layers of V1 to V4 is not shown.

as well as connections between pathways that are found to exist at all levels. Many of these strategies are used repetitively throughout the cortex, from which we derive some principles regarding the functional logic of cortical connections and make predictions about the general functional organization of uncharted cortical areas. Ultimately our hope is to provide an anatomical basis for tackling Lashley's question.

Segregation and specialization

The division of the visual system into specialized pathways begins at the very earliest stage, the retina, and continues through the subcortical visual relay, the lateral geniculate nucleus. In the retina, there are two major classes of ganglion cells, the $P\alpha$ and $P\beta$ cells^{10,11}. These have different properties and project separately to the cortex through the lateral geniculate nucleus— $P\alpha$ cells through its magnocellular (M) layers and $P\beta$ cells through its parvocellular (P) layers. Retinal ganglion cells and lateral geniculate cells have similar properties. Hence the two pathways are known as the P and M systems, terms that are

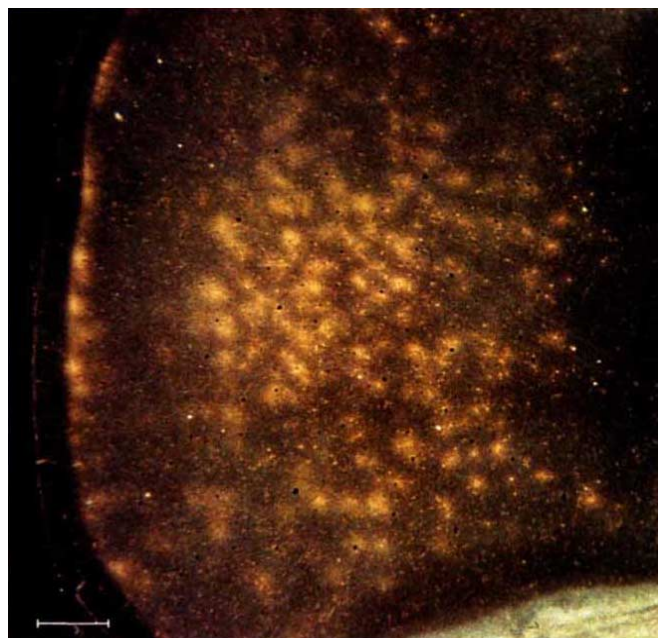
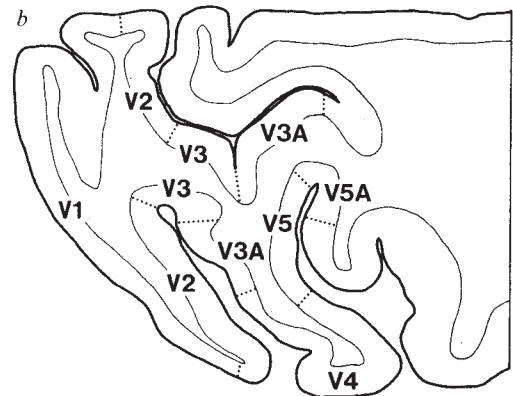
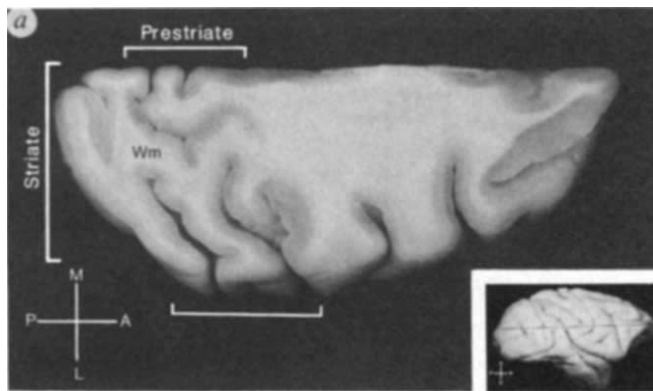


Fig. 3 The patchy connections of the primary visual cortex (area V1) with one of the specialized areas of the prestriate cortex, area V5, revealed by injecting V5 with the enzyme horseradish peroxidase. The enzyme is taken up by the terminals in the area and transported back to the cell bodies in V1; these appear golden. The section was cut parallel to the cortical surface, through the layer of V1 that contains the cells projecting to V5. The patchiness is evidence that not all cells in V1 have direct connections with V5, indicative of further functional subdivision in V1. Scale bar, 100 μm .

applied throughout the visual system because they remain segregated in and beyond the primary visual cortex. Throughout, the two systems are involved in processing different attributes of the visual scene, building on the basic characteristics already evident in the retina and lateral geniculate nucleus. $P\beta$ and P cells are selective for wavelength and have slow tonic responses, whereas $P\alpha$ and M cells are not wavelength selective, have faster, transient responses, larger receptive fields and greater sensitivity to contrasts. They also have faster conduction velocities¹²⁻¹⁴.

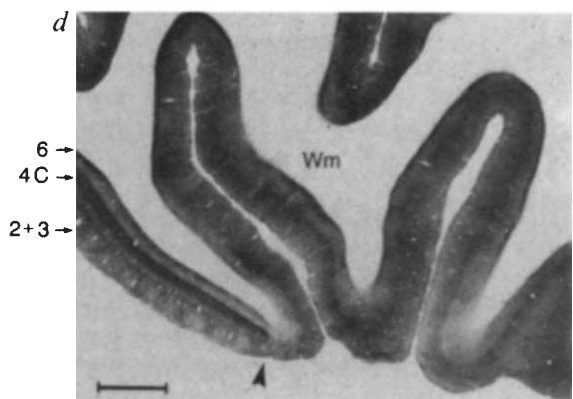
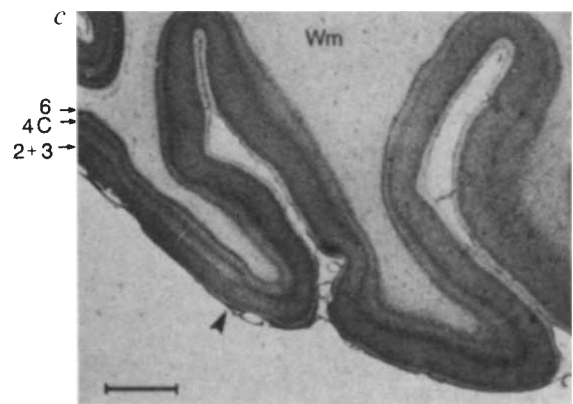
The P and M systems have different destinations in the visual cortex (Box and Fig. 1). They project to different parts of V1^{15,16}, the primary visual area, and of the second visual area (V2)^{17,18} which surrounds it. V1 and V2 have a distinctive modular organization which is fortuitously revealed by staining for the metabolic enzyme, cytochrome oxidase. V1 contains darkly staining columns perpendicular to the cortical surface and intersecting the layers. The columns are especially prominent in layers 2 and 3 where they form 'blobs', separated from each other by the more lightly staining regions termed interblobs^{19,20}. In V2, there are alternate thick and thin dark stripes, separated from each other by more lightly staining interstripes^{21,22} (Fig. 2). Both V1 and V2 project to more specialized visual areas, including V3, V4 and V5 (also termed MT)^{17,18,23}.

P-cell outputs are relayed to the blobs and interblobs of layers 2 and 3 of V1²⁴. Wavelength selectivity is concentrated in the blobs, in cells which lack orientation selectivity; many have double-opponent concentric receptive fields different from those found in lateral geniculate cells²⁵ and are especially suited to extracting information about colour. By contrast, interblob cells are usually orientation selective²⁵ and are capable of responding to contours generated by differences in wavelength or luminance^{26,27}. Mathematical considerations suggest that P cells signal a combination of colour and luminant variations in the visual field, the balance depending on the sizes, or spatial frequencies,



The visual cortex

Both in man and macaque monkey, shown here, the cerebral cortex is a highly convoluted sheet, much of it buried inside the folds or sulci, seen in (a), a section through the brain at the level indicated in the inset in (a). Visual cortex occupies the posterior third of the cortex and can be divided into many distinct visual areas, shown in (b) on a drawing of a section at the same level as in (a). One of the oldest methods used to identify areas anatomically is the study of cytoarchitecture, the grouping of cells into different layers. A section through the posterior third of the macaque brain treated to stain the cell bodies (c) shows two distinctive regions, the primary striate cortex, area V1 (left) and the prestriate cortex lying in front of it. The border is indicated by an arrowhead. The striate cortex, so named because of its richly laminated appearance, receives all the fibres from the retina through the lateral geniculate nucleus, a relay nucleus below the cortex. The prestriate cortex has a less elaborate cytoarchitecture. Staining for the activity of the metabolic enzyme cytochrome oxidase reveals complex subdivisions in the striate and prestriate areas (d), shown in more detail in Fig. 2. Techniques for tracing anatomical connections between the striate and prestriate areas and functional studies reveal several distinct areas in the prestriate cortex (V2–V5) (b), each specialized to process different attributes of the visual scene. Wm, white matter.



of the patterns to which they respond²⁸. The signal thus contains primarily wavelength information at low spatial frequencies, to which blob cells are more sensitive²², and luminance information at high spatial frequencies, to which interblob cells are more sensitive²².

Thus the P system seems to split in the upper layers of V1 into one pathway concerned primarily with colour and another primarily with form. But both lead through V2 to the same area, V4, the blobs by way of the thin stripes, and the interblobs by way of the interstripes^{25,29,30}, subdivisions of V2 which maintain these functional distinctions^{29–31}. As a whole, V4 is specialized for colour^{1,32}. Whether form signals are processed independently of colour in V4 and whether the two subdivisions of the P system retain independent identities within it is not known^{33–35}.

The M system input terminates quite separately, in layer 4B of V1²⁴ where, again, at least two new categories, orientation and direction selectivity, are generated^{36,37}. The M pathway then projects to areas V3 and V5, both directly^{38,39} and through the thick stripes of V2^{29,30}. Like layer 4B of V1, the thick stripes of

V2 are characterized by a concentration of directionally selective cells and an absence of wavelength selectivity^{29–31}. Most cells in V5 are also directionally selective—to such an extent that V5 was called the motion area when first described⁴⁰. By contrast, the chief feature of V3 is orientation selectivity⁴¹.

Separation of motion, form and colour

The temporal properties of cells in the M system give it a much greater ability to follow events changing with time, consistent with its involvement in the analysis of motion, as seen especially in V5^{40,42,43}. Orientation selectivity is generated in both the P and the M systems so it seems that both are involved in form perception, the P system through its inputs to V4 and the M system through V3. But most V4 cells, unlike V3 cells, are also wavelength selective, and the nature of the relationship between colour and form processing in V4 remains to be determined. It seems that the M system is relatively more concerned with the form of moving objects and with generating structure from motion, whereas the smaller fields and larger number of cells,

at least within the lateral geniculate nucleus and V1, enables the P system to undertake a finer grain construction of a static visual image that incorporates both wavelength and luminance information. The restriction of motion and colour to separate cortical areas explains clinical observations of cerebral lesions leading to pure losses of these attributes^{5,6}, whereas form vision, which involves both the M and P systems, is normally less impaired.

The anatomy of segregation

We can now examine some of the general principles governing cortical connections. The cortex is a highly convoluted, thin and layered sheet of tissue whose surface area greatly exceeds its thickness (see Box). Variations in its structure, which correspond to the segregation of function, are expressed along all three dimensions and no single anatomical method can reveal them all. Conventional techniques for examining the architecture of the cortex—the manner in which cells and myelinated fibres are grouped together—have defined the layering in depth of the sheet. They have generally been less successful at revealing structural variations across its length and breadth, that is, within a single area or between areas. Consequently this aspect of cortical organization was overlooked in the past, even when some functional evidence was available, until new anatomical techniques revealed the subdivisions. The two examples we mention below illustrate the inferences about functional segregation that can be drawn from anatomical observations. First, however, we describe the segregation of function across the more prominent layering of the cortex. This can be related systematically to the hierarchical organization of pathways.

Laminar segregation of function. By convention the cortex is divided into six layers, with layer 1 at its surface and layer 6 at the junction with white matter (see Box). V1 is unusual in having a more complex pattern of layering with cells related to different visual attributes segregated between its layers. But more generally throughout the cortex cells in different layers vary in their functions by playing different roles in the input–output relationships of an area (Fig. 4*a–d*). Forward projections, for instance the projection from V2 to V5, originate predominantly from the upper cortical layers and terminate in layer 4. They are invariably reciprocated by an asymmetrically organized backward projection, for example from V5 back to V2, that arises from both the upper and lower layers, but its termination avoids layer 4 to concentrate in layers 1 and 6^{44–47}. Thus the terms ‘forward’ and ‘backward’, in addition to describing the direction of a connection within a hierarchical sequence, also correspond to a particular pattern of laminar organization, although such a pattern has not yet been shown to be universal for the visual cortex.

Because the forward and backward outputs of an area arise from segregated populations of cells, it seems likely that they transmit non-identical signals. This, in a sense, is a form of specialization different from that which we have so far considered. But it is one about which there is little physiological knowledge. There are also reciprocal connections between a pair of areas that arise and terminate equally in all layers, giving a symmetrical organization intermediate between the forward and backward patterns. These connections are referred to as ‘sideways’ or ‘lateral’.

Divergent connections as an indication of functional segregation. It has been axiomatic in cortical studies that functional differences are reflected in anatomical ones and, conversely, that an area of uniform structure should also be one of uniform function⁴⁸. Because the early studies of cortical architecture resulted in a rather uniform picture, the existing clinical evidence which suggested that a part of the human prestriate cortex is specifically involved in colour vision⁴⁹ was actually dismissed, largely because the area involved could not be distinguished from the rest of the prestriate cortex⁵⁰.

The first evidence for the separation of form, motion and colour in the visual system¹ came from the identification of

distinct areas, V2, V3, V4 and V5, in the architecturally uniform prestriate cortex because each receives separate inputs from V1 and from the opposite hemisphere^{17,18,51}. The discovery of independent outputs from each mm² of V1 to separate prestriate areas provided a powerful hint for the functional specialization of these areas³³, later confirmed by physiological studies¹. Because each mm² of V1, which represents a particular region of visual space, was known to contain a functionally heterogeneous population of cells³⁷, it seemed likely that V1 would send different rather than identical categories of signals to each prestriate area. In other words, it must act as a segregator of different types of signal³³, a role also played by V2 for similar reasons.

Modular architecture and stripy connectivity as an indication of functional segregation in an area. From this evidence it was reasoned³³ that V1 and V2 might themselves conceal a separation of form, motion and colour. Such a prediction was inconsistent with the then accepted picture of V1, which was heavily influenced by the work of Hubel and Wiesel, who had reported all cells outside layer 4C to be orientation selective^{36,52}. And that work in turn was influenced, to some degree at least, by the apparent architectural uniformity of V1: for “with no anatomical indication of nonhomogeneity in the upper cortical layers, it would have been easy to dismiss occasional, apparently sporadic groups of unoriented cells”²⁵. Hubel and Wiesel’s picture of V1, then, contained no hint of the separation of form, motion and colour—a concept which has now come to dominate our thinking about the visual system—even though such a separation had already been demonstrated in the prestriate visual cortex¹.

In a repeat of history, the evidence for a separation of function in V1, as well as V2, was pursued only after non-uniformity had been shown anatomically—this time the modular or patchy architecture revealed by cytochrome oxidase histochemistry. This led to two correlations: that the specialized groupings of cells in V1 and V2 either avoid or coincide with the blobs and stripes, and that the pattern of outputs from either area to the specialized areas of the prestriate cortex, and to each other, are also patchy, the output patches again either avoiding or coinciding with the cytochrome oxidase patches. The latter finding was particularly significant because patchy patterns of connection are now known to be a widespread feature of cortical connectivity.

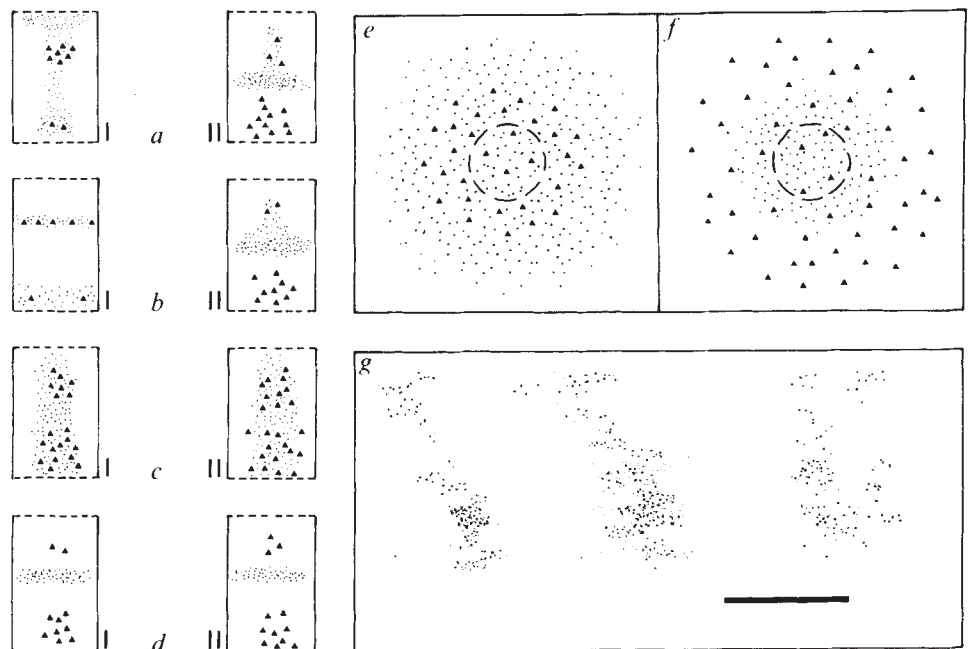
The inferences of anatomy. The lesson to be learned is that if an area has divergent projections, a modular architecture, or stripy connections with one or more areas, it is also likely to contain functionally distinct, and segregated, groups of cells. If any one of these features is observed it may be inferred that the others are also present. Layer 4B of V1 is a good example. It is known to project to both V3 and V5^{38,39} and we have recently found that its connections with V5 arise from a fairly regular array of patches (Fig. 3), similar in appearance to the blobs though not systematically related to them. This leads us to suppose that there is a further system of segregation in V1 and that cells in these patches differ from those outside, perhaps in the finer details of the analysis of motion.

Table 1 Examples of the two types of convergence

	Topical	Confluent
Forward	Blobs → thin stripes → V4 V1 → V5	V2 (thin stripes) + V3 → V4 V4 + V5 → IPS + STS
Intrinsic	in V4	in V1, V2 and IPS
Lateral		V4 ↔ V5
Backward	V5 → V1	V3 + V5 → V1 (layer 4B) V4 + V5 → V2 (K + I + N stripes)

For abbreviations see Fig. 1.

Fig. 4 Schematic representations of asymmetries in the connections between cortical areas, where projecting cells from an area are shown as pyramids and the fibres terminating in an area are shown in stippling. *a-d* show asymmetries in laminar patterns of connection. *a*, The standard pattern in which area I sends a forward projection to area II and II a backward projection to I. *b*, A variant, the connections between V1 (I) and V5 (II). It is atypical because in V1 both the cells of origin of the forward projection and the terminals of the backward projection are in the same layers. *c*, Symmetrical organization of most other connections that are not readily identifiable as either forward or backward. Many other symmetrical patterns, such as *d*, could be envisaged but have never been observed. *e*, *f*, Topical convergence in the asymmetrical connections between two areas: cells and terminals in *e* are the source of output to and the target of input from a circumscribed unit in *f* indicated by a dashed outline, and vice versa. The forward projection to a small part of *f* from a larger zone in *e* is characteristic of topical convergence. The asymmetry demonstrated by the different extents of the cells and terminals in *e* and *f*—some parts of *e* receive input from parts of *f* to which they do not project and some parts of *f* send output to parts of *e* from which they do not receive input—provides further topical convergence in the back projection from *f* to *e*. *g*, An experimental reconstruction of asymmetry similar to that in *e*, showing a small section of V2 and its connections with V5. Whereas the cells in V2 (heavy dots) projecting to V5 are grouped together in stripes, the return input to V2 from V5 (stipple) is more widely distributed and includes the territory between the stripes giving a pattern of confluent convergence.



Correlations between these features may prove to be a ubiquitous phenomenon in the organization of the cerebral cortex. All areas so far investigated make connections with many others and most of these are patchy in both origin and termination. The prediction is that they will also be found to possess a modular architecture and functional segregation. Perhaps the most studied visual area other than V1 and V2 is V5, which we also predict has an internal subdivision of function. The input to V5 from V1 and V2 terminates patchily^{53,54}, as does the input from V5 in the opposite hemisphere⁵⁵. In addition V5 is itself the source of widespread divergent projections^{47,56,57} and, at least in some species, it has a lattice-patterned cytochrome oxidase architecture^{58,59}. The functional logic of cortical connections therefore predicts that V5 is also a segregator, with functionally distinct groups of cells occupying anatomically distinct subdivisions.

Functional integration

We use the term integration broadly, to cover a number of distinct operations performed by the visual system. Perhaps the simplest is that envisaged by hierarchical theories of brain function, the progressive elaboration of increasingly more complex response properties in a specialized pathway by a pattern of spatially convergent connections repeated at each stage. A classic example is the simple-complex-hypercomplex model^{9,52} for the progressive elaboration of the orientation, position and length of a contour.

A second type of integration concerns the combination or interaction of attributes between pathways. Previous discussions have emphasized the divergence between the M and P pathways, or 'streams', and their continuation beyond the prestriate cortex—the M pathway, concerned chiefly with motion/spatial awareness, leading into the parietal lobe and the P pathway, concerned with object recognition, leading into the temporal lobe⁶⁰⁻⁶². The question as to how categories within these broad subdivisions are associated has been ignored or deferred to a further structure such as frontal cortex or the limbic system⁶¹. But the notion of two isolated, hierarchically organized and

functionally specialized pathways fails to explain one of the single most impressive facts about cortical connections—that no area of the cerebral cortex connects with only one or two other areas. Instead, from the multiple divergent projections of all areas we can deduce that the visual system employs an intricate combination of both hierarchical and parallel strategies, endowing it with a richness of connectivity that must facilitate all manner of interactions between different specializations. This is in accord with the subtle interplay between attributes known to occur perceptually. For instance velocity gradients can give information about relative depths, but erroneous depth perception can also give rise to an illusory percept of motion (the Mach illusion)⁶³.

This is an instance of one attribute being used in the computation of another but some attributes, of which motion and colour are the best examples, are probably computationally entirely independent of each other. But the percepts of motion and of colour must both be referred to the correct object in the field of view which is their source. This, third, kind of integration might use a different mechanism but, to indicate that the activity arising separately in each area is due to the same object, it too must involve intercommunication between the specialized pathways.

Integration through convergence

The anatomical means of assembling information is convergence. We define two categories, topical and confluent convergence. The first operates within a specialized pathway and involves integration across space; the second operates between pathways and involves integration between different attributes. Both are found in all types of cortical projections, forward, backward and lateral or intrinsic (Table 1), providing a substrate for convergent associations between the specialized pathways at many levels, including those at which functional segregation is first established. This leads us to a theory of multi-stage integration in the visual system.

Forward and lateral connections. Forward topical convergence occurs wherever a small part of a higher area receives input

from a larger part of a lower area (Fig. 4e,f). This generates a set of cells at each stage that have larger receptive fields and more complex properties than those at the previous stage. In the colour pathway, for instance, many blobs in V1 project to one thin stripe in V2²⁵ and several thin stripes to a small (~4 mm²) part of V4^{29,30}, such that cells in V4 have larger receptive fields and more complex properties than their counterparts in V1.

The strategy of generating larger receptive fields—that is, gathering information from a wider part of the field of view—is probably essential to the computational process. In colour vision, for example, the property of colour constancy—colours of surfaces are largely unaffected by the spectral composition of the prevailing illumination—depends on a comparison of the spectral composition of the light reflected from many different surfaces in the field of view⁶⁴. Some V4 cells have the property of colour constancy, which is absent from V1 cells⁶⁵. This increase in complexity results from the comparison of information from larger parts of the field of view, to which the widespread intrinsic connections of V4 may also contribute⁶⁵.

Confluent convergence includes any pattern of connectivity where one area receives input from several others. It is commonly considered to be the means of uniting signals from different sources, first within a single modality such as vision, and then between modalities, for example, vision and audition⁶⁶. In the visual system (Fig. 1), confluent convergence occurs both within a pathway (for instance, in the M pathway the convergence of the output of areas V1, V2 and V3 to V5) and also between the different functional pathways (shown by heavier arrows in Fig. 1). The latter involves not only the direct connections that V3, V5 and V4 make with each other but also the connections that each makes with higher areas in temporal and parietal cortex.

V4, like V5, receives a forward projection from V3, but the interconnections between V4 and V5 are of the lateral variety^{47,56}. The direct projections of V4 and V5 to the temporal lobe are largely to separate zones^{47,56,67}; but these zones are known to be interconnected⁶⁸ and physiological evidence for convergence has been obtained in a region where neurons selective for faces are found in juxtaposition with neurons which respond to characteristic locomotor or other body movements⁶⁹. The former is likely to represent the culmination of a pathway derived from the P system through V4, but possibly also involving V3, and the latter the culmination of a pathway derived from the M system through V5. In the parietal lobe the inputs from V4 and V5 converge directly at the same region^{47,55,56,70}, whereas that from V3 (through V3A, a neighbouring area with similar properties) is predominantly to a separate zone (S.Z. and S.S., unpublished results). But, the two zones are probably also interconnected (S.Z. and S.S., unpublished results) and both project into the temporal cortex^{68,71}.

The fact that both V4 and V5 have projections to the parietal as well as the temporal cortex is good evidence that the proposed division into two isolated and serially organized 'streams'^{60–62} is an oversimplification. Without denying the evident specialization of the temporal and parietal lobes, the convergent anatomy emphasizes the scope that exists for interactions between them. It seems that an area performing a specialized higher function will tap any source of information that is useful. V3 provides a forward output to both of these areas but, because of its independent output to parietal cortex, it should not be thought of simply as a step in these pathways. Instead, like V4 and V5, we regard it as a pivotal area for marshalling and redistributing the output of V1 and V2 to higher areas.

Intrinsic connections. The projections from two areas may both lead to a third area, but terminate in different parts or layers within it. Intrinsic connections between their termination zones provides another type of convergence. The projections to the temporal and parietal lobes from the specialized prestriate areas are stripy (S.Z. and S.S., unpublished results) though it is not known how the inputs from different areas relate to each other.

But even if they occupy non-overlapping zones, they would still fall within the range of lateral intrinsic connections that would permit interactions between them. An obvious analogy is in V1, where the inputs from the two eyes maintain separate territories in layer 4C, but converge to yield binocularly driven cells thereafter⁵². A similar process could integrate the information from V4 and V5 in the intraparietal sulcus (see Fig. 1), if the patchy inputs from these two areas do not coincide directly.

Intrinsic connections can also provide communication between the specialized pathways at earlier levels. The horizontally coursing fibre plexus in layer 3B of V2 contains collateral axons which are sufficiently long (up to 4 mm⁷²) to allow interactions between all three types of stripe. And in V1 there are modest connections between layers 2 and 3 and layer 4B⁷³. No sign of interaction has been detected physiologically, so it seems that these connections are not strong enough to undermine the essentially segregatory processes in these areas; hence their particular influence remains to be understood.

Backward connections. Most connections between areas in the cortex appear to be reciprocal but within an area the distributions of the forward and backward components do not precisely coincide (Fig. 4). This asymmetry gives rise to confluent and topical patterns of convergence in the backward direction. Good examples are the projections of V5 and V4 back to V2⁵⁹ (Fig. 4g). V5 projects back most heavily to the thick stripes from which it receives its input but the projection also encompasses the territory of the intervening thin stripes and interstripes. Similarly, the back projection from V4 is most concentrated in the thin stripes and interstripes but also stretches across the thick stripes. Thus areas V4 and V5 have confluent projections back to each type of stripe in V2, so that they may influence each other at an earlier stage in their respective pathways than through the direct link between them. A similarly asymmetrical backward system is found in layer 4B of V1, which contains cells projecting to both V3 and V5. The diffuse back projection from V5 to V1 thus has the potential of influencing the cells with output to V3—and hence of informing one subdivision of the M system of the activity in another. It is interesting that the back projection from V5 to V1 is atypical in that it avoids the upper layers of V1, where the P system is situated; the backwardly-mediated interaction between the M and P systems is thus deferred until V2.

The back projections from V5 to V1 and V2 also provide a system of topical convergence, because the output from a prescribed part of V5 is sufficiently diffuse to influence regions of V1 and V2 which do not themselves provide an input to that part of V5 (Fig. 4e,f). But whereas forward topical convergence produces larger excitatory receptive fields, this is not so for backward convergence: backward connections seem not to excite cells in lower areas, but instead influence the way they respond to stimuli within their smaller receptive fields. A direction-selective cell in V1, for instance, might be less excited by its preferred stimulus when many objects outside its classical receptive field are also moving in the same direction⁷⁴.

Multi-stage integration in the visual cortex

The functional logic of the cortical connections described above is twofold—to achieve segregation and to allow for integration. The first two visual areas, V1 and V2, are segregators, containing separate groupings of orientation, wavelength, and direction selective cells. Divergent and convergent outputs from V1 and V2 assemble signals related to these basic attributes of vision in three pivotal areas: areas V3 and V3A (the V3 complex) involved in dynamic form analysis; area V4 (the V4 complex) involved with colour in association with static form analysis; and the V5 complex (areas V5 and V5A) involved with motion. None of these functions is necessarily unique to these areas, as implied by the term specialization, because each is part of a pathway that involves several areas. Furthermore, cells selective for certain basic attributes, such as disparity, may be found in

each⁷⁵. The pivotal areas may communicate with each other at three distinct levels: directly; through confluent backward connections with V1 and V2; and through confluent forward connections with areas in the parietal and temporal lobes. We suppose that these higher areas also communicate with each other using the same three strategies, further emphasizing the multi-stage integrative process used to produce the visual image in the brain.

Because the patterns of connectivity we describe here are found throughout the cortex, we further suppose that similar strategies for segregation of signals and their subsequent integration are common to the operations of all parts of the cerebral cortex.

The authors' work reported here was supported by grants from the Wellcome Trust and the SERC.

1. Zeki, S. M. *Nature* **274**, 423-428 (1978).
2. Newsome, W. T., Wurtz, R. H., Dürsteler, M. R. & Mikami, A. J. *J. Neurosci.* **5**, 835-840 (1985).
3. Wild, H. M., Butler, D., Carden, D. & Kulikowski, J. J. *Nature* **313**, 133-135 (1985).
4. Heywood, C. A. & Cowey, A. J. *J. Neurosci.* **7**, 2601-2617 (1987).
5. Damasio, A., Yamada, T., Damasio, H., Corbett, J. & McKee, J. *Neurology* **30**, 1064-1071 (1980).
6. Zihl, J., von Cramon, D. & Mai, N. *Brain* **106**, 313-340 (1983).
7. Lashley, K. S. *Science* **73**, 245-254 (1931).
8. Hochberg, J. E. *Perception* (Prentice Hall, New Jersey, 1964).
9. Hubel, D. H. & Wiesel, T. N. *J. Neurophysiol.* **28**, 229-289 (1965).
10. Leventhal, A. G., Rodieck, R. W. & Dreher, B. *Science* **213**, 1139-1142 (1981).
11. Perry, V. H., Oehler, R. & Cowey, A. *J. Neurosci.* **12**, 1101-1123 (1984).
12. Wiesel, T. N. & Hubel, D. H. *J. Neurophysiol.* **29**, 1115-1156 (1966).
13. Dreher, B., Fukada, Y. & Rodieck, R. W. *J. Physiol.* **258**, 433-452 (1976).
14. Schiller, P. H. & Malpeli, J. G. *J. Neurophysiol.* **41**, 788-797 (1978).
15. Hubel, D. H. & Wiesel, T. N. *J. comp. Neurol.* **146**, 421-450 (1972).
16. Hendrickson, A. E., Wilson, J. R. & Ogren, M. P. *J. comp. Neurol.* **182**, 123-136 (1978).
17. Cragg, B. G. *Vision Res.* **9**, 733-747 (1969).
18. Zeki, S. M. *Brain Res.* **19**, 63-75 (1969).
19. Horton, J. C. & Hubel, D. H. *Nature* **292**, 762-764 (1981).
20. Horton, J. C. *Phil. Trans. R. Soc. B.* **304**, 199-253 (1984).
21. Livingstone, M. S. & Hubel, D. H. *Proc. natn. Acad. Sci. U.S.A.* **79**, 6098-6101 (1982).
22. Tootell, R. B. H., Silverman, M. S., DeValois, R. L. & Jacobs, G. H. *Science* **220**, 737-739 (1983).
23. Zeki, S. M. *Brain Res.* **34**, 19-35 (1971).
24. Lund, J. S. & Boothe, R. G. *J. comp. Neurol.* **159**, 305-334 (1975).
25. Hubel, D. H. & Livingstone, M. S. *J. Neurosci.* **4**, 309-356 (1984).
26. Gouras, P. & Kruger, J. *J. Neurophysiol.* **42**, 850-860 (1979).
27. Thorell, L. G., DeValois, R. L. & Albrecht, D. G. *Vision Res.* **24**, 715-769 (1984).
28. Ingling, R. & Martinez-Urriegas, E. *Vision Res.* **23**, 1495-1500 (1983).
29. Shipp, S. & Zeki, S. *Nature* **315**, 322-325 (1985).
30. DeYoe, E. A. & Van Essen, D. C. *Nature* **317**, 58-61 (1985).
31. Hubel, D. H. & Livingstone, M. S. *Nature* **315**, 325-327 (1985).
32. Zeki, S. M. *Brain Res.* **53**, 422-427 (1973).
33. Zeki, S. M. *Cold Spring Harb. Symp. quant. Biol.* **16**, 591-600 (1975).
34. Zeki, S. *Proc. R. Soc. B.* **217**, 449-470 (1983).
35. Desimone, R. & Schein, S. J. *J. Neurophysiol.* **57**, 835-868 (1987).
36. Hubel, D. H. & Wiesel, T. N. *J. Physiol., Lond.* **195**, 215-243 (1968).
37. Dow, B. M. *J. Neurophysiol.* **37**, 927-946 (1974).
38. Burkhalter, A., Felleman, D. J., Newsome, W. T. & Van Essen, D. C. *Vision Res.* **26**, 63-80 (1986).
39. Lund, J. S., Lund, R. D., Hendrickson, A. E., Bunt, A. H. & Fuchs, A. F. *J. comp. Neurol.* **164**, 287-304 (1975).
40. Zeki, S. M. *J. Physiol., Lond.* **236**, 549-573 (1974).
41. Zeki, S. M. *J. Physiol., Lond.* **277**, 273-290 (1978).
42. Maunsell, J. H. R. & Van Essen, D. C. *J. Neurophysiol.* **49**, 1127-1147 (1983).
43. Albright, T. D. *J. Neurophysiol.* **52**, 1106-1130 (1984).
44. Gilbert, C. D. & Kelly, J. P. *J. comp. Neurol.* **163**, 81-106 (1975).
45. Tigges, J., Tigges, M. & Perachio, A. A. *J. comp. Neurol.* **176**, 87-100 (1977).
46. Rockland, K. S. & Pandya, D. N. *Brain Res.* **179**, 3-20 (1979).
47. Maunsell, J. H. R. & Van Essen, D. C. *J. Neurosci.* **3**, 2563-2586 (1983).
48. Lashley, K. S. & Clark, G. J. *comp. Neurol.* **85**, 223-305 (1946).
49. Economo, C. von & Koskinas, G. N. *Die cytoarchitektonik der Grosshirnrinde des erwachsenen Menschen* (Springer, Berlin, 1925).
50. Lashley, K. S. *Genet. Psychol. Monogr.* **37**, 107-166 (1948).
51. Van Essen, D. C. & Zeki, S. M. *J. Physiol.* **277**, 193-226 (1978).
52. Hubel, D. H. & Wiesel, T. N. *Proc. R. Soc. B.* **198**, 1-59 (1977).
53. Montero, V. M. *J. comp. Neurol.* **189**, 45-59 (1980).
54. Zeki, S. M. *Proc. R. Soc. B.* **193**, 199-207 (1976).
55. Zeki, S. M. *Proc. R. Soc. B.* **197**, 195-223 (1977).
56. Ungerleider, L. G. & Desimone, R. *J. comp. Neurol.* **248**, 190-222 (1986).
57. Zeki, S. in *Coding and Efficiency in the Visual System* (ed Blakemore, C.) (Cambridge University Press, 1988).
58. Tootell, R. B. H., Hamilton, S. L. & Silverman, M. S. *J. Neurosci.* **5**, 2786-2800 (1985).
59. Shipp, S. & Zeki, S. *Eur. J. Neurosci.* (submitted).
60. Van Essen, D. C. & Maunsell, J. H. R. *Trends Neurosci.* **6**, 370-375 (1983).
61. Mishkin, M., Ungerleider, L. G. & Macko, K. A. *Trends Neurosci.* **6**, 414-417 (1983).
62. Desimone, R., Schein, S. J., Moran, J. & Ungerleider, L. G. *Vision Res.* **25**, 441-452 (1985).
63. Ullman, S. *The Interpretation of Visual Motion* (MIT, Cambridge, Mass., 1979).
64. Land, E. H. *Proc. R. Inst. Gt Br.* **47**, 23-57 (1974).
65. Zeki, S. *Neurosci.* **9**, 741-765 (1983).
66. Jones, E. G. & Powell, T. P. S. *Brain* **93**, 793-820 (1970).
67. Desimone, R., Fleming, J. & Gross, C. G. *Brain Res.* **184**, 41-55 (1980).
68. Seltzer, B. & Pandya, D. N. *Brain Res.* **149**, 1-24 (1978).
69. Perret, D. J. *et al. Behav. Brain Res.* **16**, 153-170 (1985).
70. Seltzer, B. & Pandya, D. N. *Brain Res.* **192**, 339-351 (1980).
71. Seltzer, B. & Pandya, D. *Expl Brain Res.* **55**, 301-312 (1984).
72. Rockland, K. S. *J. comp. Neurol.* **235**, 467-478 (1985).
73. Blasdel, G. G., Lund, J. S. & Fitzpatrick, D. *J. Neurosci.* **5**, 3350-3369 (1985).
74. Allman, J. & McGuinness, E. in *Signal and Sense: Local and Global Order in Perceptual Maps* (eds Edelman, G., Gall, E. & Cowan, W. M.) (Wiley, New York, 1989).
75. Zeki, S. *Proc. R. Soc. B.* **204**, 379-397 (1979).

ARTICLES

Controls of the structure of subducted slabs

Michael Gurnis* & Bradford H. Hager

Seismological Laboratory, California Institute of Technology, Pasadena, California 91125, USA

Numerical simulations of subducting slabs are formulated in which the shape and dip of the slab are determined by the dynamics of the flow, rather than imposed a priori. The dip of slabs is a function of the time since the initiation of subduction. Slabs fold, develop a kink in dip, and thicken on entry into a high-viscosity lower mantle. Comparison of the simulations with seismic observations suggest that the lower mantle is at least 10-30 times more viscous than the upper mantle.

THE shapes and dips of seismic Benioff zones are the strongest constraint on the pattern of flow in the mantle. Benioff zones always have positive dips and generally lack any major change in dip just above the 670-km discontinuity, which suggests that slabs penetrate into the lower mantle¹⁻³. This is consistent with interpretations by residual sphere analysis⁴ that there are tongues of seismically fast mantle directly under Benioff zones in the Western Pacific. These observations at first seem to be at odds with the down-dip compression of slabs that penetrate through the transition region⁵, with the contortions apparent in the Tonga

slab from seismicity⁶, and with a detailed tomographic study of Western Pacific slabs⁷ which shows that slabs are apparently deflected away from the 670-km discontinuity in a few places. Two hybrid models of mantle dynamics have been advanced⁸ in an attempt to reconcile these observations: mantle-wide flow with depth-dependent viscosity, and penetrative convection. Both models allow slab penetration into the lower mantle, at the same time having increased resistance to flow through the seismic transition zone (400 to 670 km depth); one model has a viscosity jump and the other a compositional jump at 670 km depth.

Here we explore the first hybrid model, mantle-wide flow with depth-dependent viscosity, using a new class of finite element

* Present address: Department of Geological Sciences, University of Michigan, Ann Arbor, Michigan 48109, USA.

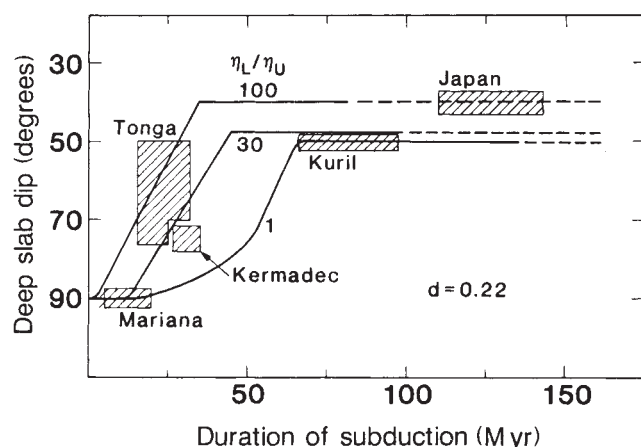


Fig. 3 Duration of subduction versus deep slab dip for both Western Pacific slabs and numerical models. For the modelled cases, the dip was measured just above the viscosity interface, as illustrated in Fig. 2c. The deep dip for observed slabs comes from the trend in seismicity in the transition region (400–670 km depth). For the Kermadec, Kuril and Japan slab, average dips are from ref. 15. The range in dips for Tonga and Mariana are standard derivations of many dip measures (from ref. 6 and H.-W. Zhou, personal communication). In general, the observed duration of subduction is determined from the oldest volcanic rock samples in a given arc; age data from ref. 25. For the Marianas, there is a significant discrepancy between this age (~40 Myr) and the age determined from the magnetic lineations (<20 Myr) (ref. 26); the latter age was used, being judged as the more reliable.

subduction (Fig. 2c), the dip just above the viscosity discontinuity is ~90°, whereas 30 Myr later (Fig. 2d) the dip is ~50°. Young slabs are always near vertical, but the dip shallows as the subduction zone ages; eventually, the deep dip approaches an asymptote of ~50°. The approach to steady state, however, is determined by the viscosity structure of the system; as shown in Fig. 3, slabs in models with higher values of η_L/η_U (which have smaller η_U as $\bar{\eta}$ is held constant) approach steady state faster than in models with smaller η_L/η_U . The case with the isoviscous background was the slowest to respond and did not start shallowing until the slab reached the bottom of the box.

The steady-state dip is independent of η_L/η_U and is controlled by a balance between the pressure gradient generated by the oceanic plate, which tends to lift the plate, and the buoyancy force of the slab, which tends to pull the plate toward the vertical^{11,12}. The ~50° dip found here is similar to the 63° dip determined by Stevenson and Turner¹¹ from a simplified model. The primary reasons for the shallower dip are probably the existence of trench migration (which was not allowed in the earlier models) and periodic boundary conditions. The trenches migrate 'oceanward' with respect to the centre of mass of the system at ~10% of the total convergence velocity of the continental plate-oceanic plate. The earlier calculations also differ

in being infinite in both depth and lateral extent and in having slabs that were not permitted to bend, except at the trench.

Realistic positive dips were not obtained when periodic boundary conditions were substituted at the sides. In Fig. 2b and d, all the basic material parameters are the same (for example, $\eta_L/\eta_U = 30$, $d = 0.5$), as is the duration of subduction, as indicated by the distance of ridge migration, but the boundary conditions are different: Fig. 2b has reflecting, and Fig. 2d periodic, boundary conditions; a vertical dip is obtained for reflecting conditions and a 50° dip is obtained for periodic conditions. This result is somewhat different from the experimental results of Kincad and Olson¹⁰, where positive dips were obtained when side walls were present; we suspect the primary cause of this difference was the existence of back-arc spreading in their models, a phenomenon which was not allowed in the present calculations.

Stress in slabs. The deviatoric stress in the slabs just before penetration into the lower layer is a function of the viscosity structure. As has been calculated before, slabs penetrating through a boundary with no change in viscosity are in tension above the boundary¹³ (Fig. 2a), whereas slabs penetrating into a layer with a jump in viscosity are in compression¹⁴ (Fig. 2d).

Folding. The shortest timescale phenomenon displayed was the folding of the slab just before penetration into the more viscous lower layer. Folding was only displayed for cases with $\eta_L/\eta_U > 10$. Three instants in time are shown in Fig. 4 for $\eta_L/\eta_U = 30$; at 50 Myr after initiation of subduction, a distinct fold develops, with its apex at the viscosity discontinuity (Fig. 4a). Time-dependence is indicated by the disappearance and reappearance of the fold (Fig. 4b and c). Folding did not set in until a large cold thermal anomaly had built up in the lower layer. By viewing only the slab in the upper layer (for example by covering the lower layer, see Fig. 4c) it appears as if the slab is bent away from the lower layer so that it is prevented from penetrating into the lower layer.

Slab structure in the lower layer. The most prominent change that the slab experiences after passage through the viscosity discontinuity is an advective thickening caused by the decreased slab velocity in the lower layer. This phenomenon is evident in Figs 2 and 4. For $\eta_L/\eta_U = 10$, the slab thickens by ~1.3, and for $\eta_L/\eta_U = 30$, by ~1.5; for $\eta_L/\eta_U = 10^2$ and $d = 0.5$, the thickening is difficult to quantify because the slab loses its tabular shape, although the width of the cold region is >3 times the slab thickness in the upper region. Besides thickening after penetration into the lower layer, slabs also develop a substantial kink in dip (Fig. 5a), which increases linearly in proportion to $\log(\eta_L/\eta_U)$ (Fig. 5b). For $\eta_L/\eta_U = 3$, the slab dip increases by 10°, and for $\eta_L/\eta_U = 10^2$, by 30° to 40°. The kink in dip is probably the result of continuity of shear traction across the layer boundary, leading to greater horizontal shearing in the upper layer.

Comparison with seismic observations

Much of the phenomenology shown by the finite-element calculations has previously been elucidated, or at least been hinted at, in the Earth, some observations have been known for many years (like the deep dip of seismic Benioff zones), some had recently been made (like the three-dimensional seismic velocity structure from tomography), or newly synthesised here from old observations (the arc age versus dip relation). Besides giving an overview of relevant observations, we will show that the structure of subducted slabs can be explained by a significant jump in viscosity between the upper and lower mantle (Table 1); although the procedure used is not unique, we believe a value of η_L/η_U of 10–30 is a lower bound.

Arc age versus deep dip. Cases have been made for the control of slab dip by global mantle flow¹ and by the presence or absence of back-arc spreading¹⁵. To this list we add the possibility that the dips of slabs are also controlled by the duration of subduction. Although not previously noted in the literature, there is an

Table 1 Comparison of slab models with seismic observations

Behaviour of dynamic model	Seismic observation	η_L/η_U
Time versus dip	Arc age versus deep dip	—
Folded slabs	Shape of Benioff zone from seismicity; tomography of slabs	>10
Stress	Focal mechanisms in slabs; seismicity peak above 670 km	>10
Kink in slab dip	Kink in slab dip at 670 km from residual sphere	30–100
Advective thickening	Residual sphere; SH waveforms; surface waves	>10

approximate correlation between the duration of subduction and deep dip in the Western Pacific, shown in Fig. 3 and by the results from the modelling. Data sources are given in the caption to Fig. 3. Only slabs with seismicity extending to ~ 670 km depth have been used. The three slabs younger than 30 Myr (Tonga, Kermadec, and Mariana) have dips greater than 60° , whereas the two slabs older than 75 Myr (Japan and Kurile) have dips of less than 50° . In the Western Pacific, slabs are in the 10–100 Myr age range, which is roughly the time to approach steady-state dip. The observed scatter in deep dip may be at least in part controlled by the different ages of arcs, such that the slabs are in different stages of maturing to a steady state. As discussed above, different viscosity distributions lead to different age versus dip curves, but the observed scatter in the duration of subduction and deep dip precludes any realistic constraint on the viscosity contrast across 670 km. This scatter probably results from both global mantle flow and back-arc spreading.

Contorted slabs. Slabs are often contorted in the seismic transition zone, as noted in the Western Pacific from both seismicity and seismic tomography. Giardini and Woodhouse⁶ observed that the deepest part of the Tonga slab is displaced laterally, like the folding seen in Fig. 4a and c. From the detailed P and S wave tomographic study of the Western Pacific by Zhou and Clayton⁷, the seismically fast slabs appear in places to penetrate into the lower mantle (for example in the Kuriles), but in other regions appear to lie flat at 670 km (along part of the Izu-Bonin slab, for example). In most areas, however, the fate of the high-velocity slab is ambiguous. These observations suggest that slabs are in various stages of folding above a depth of 670 km, which is consistent with an increased resistance to flow there. A jump in viscosity at the base of the seismic transition region is one mechanism to increase resistance to flow. Folding was only found for cases with $\eta_L/\eta_U > 10$, which suggests that the viscosity contrast across the 670-km discontinuity is >10 ; these observations are also consistent with a much larger contrast.

Focal mechanisms and seismicity in slabs. Slabs which penetrate to ~ 670 km are in down-dip compression, as determined from the focal mechanisms of large earthquakes^{13,5}; this is inconsistent with slab penetration into an isoviscous mantle¹³ and requires some sort of increased resistance to slab flow. Assuming slabs penetrate into the lower mantle, our calculations and many more by Vassiliou¹⁶ require η_L/η_U to be >10 . Moreover, the focal mechanisms showing down-dip compression are consistent with the level of seismicity, which shows a distinct increase between 400 km and 670 km depth. This increase in seismicity is particularly evident for the central Tonga-Kermadec region, which shows that the level of seismicity between 400–670 km is as high as the level at 0 to 100 km depth¹⁴. This increase in seismicity shows that the total energy released by earthquakes peaks just above 670 km^{13,17}. The deep seismicity requires that slabs meet increased resistance to flow, which would be consistent with a jump in viscosity¹⁴.

Slab structure in the lower mantle. Some observations show that the structure of slabs may change following penetration into the lower mantle. Fisher *et al.*¹⁸ studied the effect of slab thickening

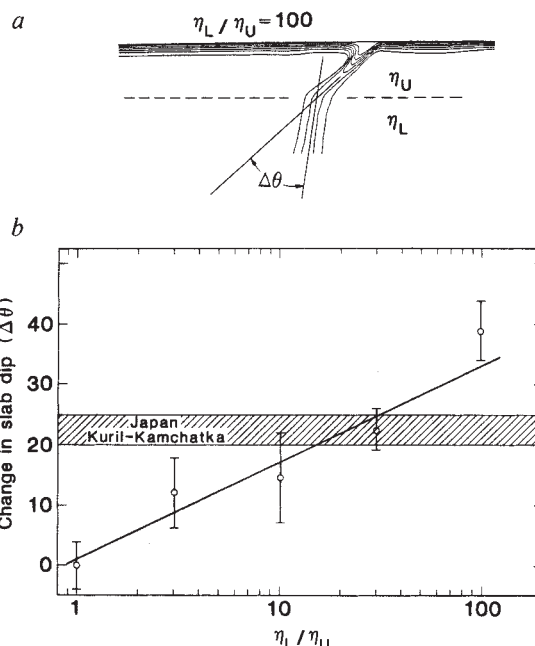


Fig. 5 a, A kink in dip of magnitude $\Delta\theta$ develops across a viscosity interface. The contour intervals are 0.2, 0.4, 0.6, 0.8 and 0.99. b, Dependence of $\Delta\theta$ on the magnitude of the jump in viscosity. Vertical error bars reflect the difference in the range in measured dips above and below the interface. The hatched region for Japan and Kuril-Kamchatka is from ref. 4.

on the travel-time anomalies of deep-focus earthquakes in slabs; they find that the travel-time data are best fitted by relatively undeformed slabs, but that the Mariana slab could be thickened by up to a factor of five. This observation is consistent with, but does not demand, a jump in viscosity of $\sim 10^2$ across the 670-km discontinuity. More robust constraints on the thickening of slabs come from the shapes of seismic waves that pass through the slabs at different angles. A narrow high-velocity slab acts as an anti-wave guide and causes waveforms to broaden¹⁹. Horizontally polarized shear waves do not broaden nearly as much as expected for a narrow slab, and suggest that the high-velocity slab widens near the 670-km discontinuity²⁰. Inversions of surface waves indicate that at a depth greater than 550 km there are broad fast regions under the mid Western Pacific and under South America which are not consistent with 100-km-thick slabs²¹. One interpretation is that slabs thicken considerably at the top of the lower mantle, which is consistent with a high-viscosity lower mantle.

Creager and Jordan⁴ found that better model fits to the travel times of slabs were obtained if the modelled slabs increase their dip in the lower mantle. The best fit to the Kurile-Kamchatka slab had a dip increasing from $\sim 45^\circ$ to $\sim 65^\circ$ and the Japan slab from $\sim 30^\circ$ to $\sim 55^\circ$. The Mariana slab, which is already vertical through the transition region, showed no change in dip. As indicated in Fig. 5, these changes in dip are not consistent with slab penetration through an isoviscous mantle and, at least for the Japan and Kurile-Kamchatka slabs, require a jump in viscosity of at least 10–30 across the 670-km discontinuity.

Conclusions

We have formulated realistic, dynamically self-consistent models of subducting slabs; the shape and dip of the slabs are determined by the flow, rather than being imposed as either boundary or initial conditions. If there is an increase in viscosity from the upper to the lower mantle, then the results of the simulations are consistent with a variety of seismic observations. In particular, slabs will be in down-dip compression, develop a kink in dip at a viscosity interface, and both fold and advect

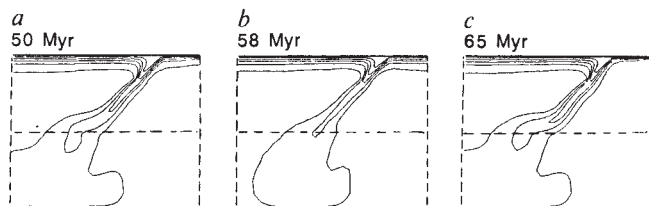


Fig. 4 Isotherms of a slab as it folds in the upper layer during penetration into the lower layer. There is a jump in viscosity by a factor of 30 across the interface (horizontal dashed line). The contour intervals are 0.2, 0.4, 0.6, 0.8 and 0.99.

tively thicken on entering a more viscous layer. These results are consistent with the focal mechanisms of earthquakes within slabs, with observed kinks in slabs from residual sphere analysis, with the shape of Benioff zones from both seismicity and tomography, and with the broadening of slabs from waveform analysis of shear waves. Taken together, these seismic observations support the hypothesis that the lower mantle is at least 10–30 times more viscous than the upper mantle, consistent with

the picture emerging from observations and models of the geoid²², from global seismic tomography²³, and from isotope geochemistry²⁴.

This work was supported by the National Science Foundation and the Ametek Corporation. Computing was carried out on the facilities of the San Diego Supercomputer Center; we thank Steve Lamont at SDSC for assistance in animating the simulations.

Received 20 April; accepted 14 July 1988.

1. Hager, B. H. & O'Connell, R. J. *J. geophys. Res.* **84**, 1031–1048 (1979).
2. Hager, B. H., O'Connell, R. J. & Raefsky, A. *Tectonophysics* **99**, 165–189 (1983).
3. Davies, G. F. *EOS, Trans. Am. geophys. U.* **67**, 381 (1986).
4. Creager, K. C. & Jordan, T. H. *J. geophys. Res.* **91**, 3573–3589 (1986).
5. Isacks, B. & Molnar, P. *Rev. Geophys.* **9**, 103–174 (1971).
6. Giardini, D. & Woodhouse, J. H. *Nature* **307**, 505–509 (1984).
7. Zhou, H.-W. & Clayton, R. W. *EOS Trans. Am. geophys. U.* **68**, 1379 (1987).
8. Silver, P. G., Carlson, R. W. & Olson, P. A. *Rev. Earth planet. Sci.* **16**, 477–541 (1988).
9. Christensen, U. R. & Yuen, D. A. *J. geophys. Res.* **89**, 4389–4402 (1984).
10. Kincaid, C. & Olson, P. *J. geophys. Res.* **92**, 13832–13840 (1987).
11. Stevenson, D. J. & Turner, J. S. *Nature* **270**, 334–336 (1987).
12. Tovish, A., Schubert, G. & Luyendyk, B. P. *J. geophys. Res.* **83**, 5892–5898 (1978).

13. Richter, F. M. *J. geophys. Res.* **84**, 6783–6995 (1979).
14. Vassiliou, M. S., Hager, B. H. & Raefsky, A. *J. Geodyn.* **1**, 11–28 (1984).
15. Uyeda, S. & Kanamori, H. *J. geophys. Res.* **84**, 1049–1061 (1979).
16. Vassiliou, M. S. thesis, Calif. Inst. Technol. (1983).
17. Abe, K. & Kanamori, H. *J. geophys. Res.* **84**, 3589–3592 (1984).
18. Fischer, K. M., Jordan, J. H. & Creager, K. C. *J. geophys. Res.* **93**, 4773–4783 (1988).
19. Vidale, J. E. *Geophys. Res. Lett.* **14**, 542–545 (1987).
20. Vidale, J. E. & Garcia-Gonzalez, D. *Geophys. Res. Lett.* **15**, 369–372 (1988).
21. Tanimoto, T. *Geophys. J.* (in the press).
22. Hager, B. H. *J. geophys. Res.* **89**, 6003–6015 (1984).
23. Hager, B. H. & Clayton, R. W. in *Mantle Convection* (ed. Peltier, W. R.) (in the press).
24. Gurnis, M. & Davies, G. F. *Geophys. Res. Lett.* **13**, 541–544 (1986).
25. Jarrard, R. D. *Rev. Geophys. Space Phys.* **24**, 217–284 (1986).
26. Karig, D. E. *Geol. Soc. Am. Bull.* **82**, 323–344 (1971).

Crystal structure of *trp* repressor/operator complex at atomic resolution

Z. Otwinowski, R. W. Schevitz*, R.-G. Zhang†, C. L. Lawson*, A. Joachimiak*, R. Q. Marmorstein‡, B. F. Luisi & P. B. Sigler§

Department of Biochemistry and Molecular Biology, and ‡ Department of Chemistry, University of Chicago, 920 East 58 Street, Chicago, Illinois 60637, USA

The crystal structure of the trp repressor/operator complex shows an extensive contact surface, including 24 direct and 6 solvent-mediated hydrogen bonds to the phosphate groups of the DNA. There are no direct hydrogen bonds or non-polar contacts to the bases that can explain the repressor's specificity for the operator sequence. Rather, the sequence seems to be recognized indirectly through its effects on the geometry of the phosphate backbone, which in turn permits the formation of a stable interface. Water-mediated polar contacts to the bases also appear to contribute part of the specificity.

HERE we describe the atomic details of a specific protein/DNA interface observed in the crystal structure of the *trp* repressor/operator complex refined to 2.4-Å resolution. The formation of this complex depends on the binding of the corepressor ligand, L-tryptophan, and represents the main mechanism for the transcriptional control of L-tryptophan levels in enteric bacteria¹. The crystal structure clearly shows the details of both direct and solvent-mediated contacts between the protein and DNA. These interactions and a substantial solvent-excluded contact surface are created by structural adjustments in both the DNA and protein which permit a far more intimate fit than anticipated from docking studies². The crystal structure also shows that, in addition to the corepressor's positive allosteric effect on the conformation of the DNA-binding domains³, the bound tryptophan itself and four of the surrounding side chains of its binding site are positioned to form hydrogen bonds to the operator to a degree not expected from modelling.

The most striking finding is that the crystal structure shows no direct hydrogen-bonded or non-polar contacts with the bases that could explain in simple chemical terms the repressor's preferential binding to the operator sequence. Instead, the direct

hydrogen-bonded contacts are mainly to the operator's phosphate groups. In each complex there are water-mediated hydrogen bonds to six bases that have been shown *in vivo*⁴ to be important in specifying the operator as the target sequence for the *trp* repressor. Specificity in the *trp* system does not rely on a simple complementarity between the repressor's amino-acid side chains and the operator's bases, but rather depends on a combination of solvent-mediated interactions with critical bases and on still poorly understood principles by which the base sequence restrains the conformation of B-form DNA.

Structure determination and refinement

The structure was determined from the monoclinic crystals described by Joachimiak *et al.*⁵; (P₂, two complexes/asymmetric unit). The crystals required about a year to grow from 35% 2,4-dimethyl pentanediol, 50 mM NaCl, 11 mM CaCl₂, 10 mM cacodylate, pH 6.8. The operator was simulated by a symmetrical 18-base-pair duplex with overhanging 5' T residues. Data were collected at room temperature with rotation photographs and area detectors. All photographic data were obtained from monochromatic synchrotron X-ray sources, either at the Stanford Synchrotron Radiation Laboratory (the parent data, $\lambda = 1.54$ Å) or at the Cornell High Energy Synchrotron Source (the PbCl₂ derivative data, $\lambda = 1.57$ Å). Synchrotron radiation was essential to obtain accurate data beyond 3.0 Å. Data from the area detector facility at the University of California, San

* Present addresses: Lilly Research Laboratories, Indianapolis, Indiana 46285, USA (R.W.S.); Department of Physical Chemistry, N. Rijksuniversiteit, Groningen, The Netherlands (C.L.L.); Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland (A.J.).

† Permanent address: Chinese Academy of Sciences, Institute of Biochemistry, Shanghai, Republic of China.

§ To whom correspondence should be addressed.

Table 1 Statistics of data collection and refinement

Resmax	Predicted*	Refined†	% Coverage‡	⟨I⟩§	⟨Error⟩	% Weak F¶	R-sym*	R-factor**	Phasing power PbCl ₂ derivative††
10.0–4.80	2,976	2,945	99.0	4,641	194	0.5	0.07	0.287	1.5
4.80–3.81	3,344	3,320	99.3	3,555	227	0.7	0.11	0.227	1.0
3.81–3.33	3,286	3,239	98.6	2,467	185	1.3	0.13	0.227	0.8
3.33–3.02	3,322	3,207	96.5	1,465	142	3.0	0.15	0.249	
3.02–2.81	3,291	3,001	91.2	583	83	7.2	0.20	0.270	
2.81–2.64	3,283	2,683	81.7	461	82	12.3	0.23	0.259	
2.64–2.51	3,296	2,448	74.3	331	78	18.3	0.29	0.243	
2.51–2.40	3,252	1,895	58.3	236	72	28.4	0.31	0.230	
Total	26,050	22,726	87.3	1,817	137	7.5	0.11	0.249	

* The number of possible reflections in each resolution shell.

† The number of reflections used in refinement.

‡ The extent of coverage of the data used in refinement, expressed as the percentage of the number predicted.

§ The average intensity, expressed in arbitrary units.

|| The average error.

¶ The number of observations with $|F|$ less than two standard deviations.

* $\sum_h \langle |I(h') - \bar{I}(h)| \rangle / \sum_h \bar{I}(h)$ where $\langle |I(h') - \bar{I}(h)| \rangle$ is the average absolute deviation of equivalent (symmetry and/or Friedel) reflections $I(h')$ from the average $\bar{I}(h)$. The largest component to R-sym was the crystal-to-crystal variation. On average, each reflection was measured 4.2 times. Eight crystals were used.

** $\sum_h ||F^o(h) - |F^c(h)|| / \sum_h |F^o(h)|$, where $|F^o|$ and $|F^c|$ are observed and calculated structure factor amplitudes respectively.

†† There were 10,595 unique reflections to 2.8-Å resolution from a single crystal giving 60% coverage. The R-sym is 5%. The r.m.s. difference between the parent and PbCl₂ structure amplitude was 28%. Phasing power of the PbCl₂ derivative, expressed as $\sum_h |F^c_H| / \sum_h [|F^o_P| \exp(i\phi) + F^c_H] - |F^o_{PH}|$, where $|F^o_P|$ and $|F^o_{PH}|$ are the observed amplitudes of the parent and heavy-atom derivative structure factors respectively, ϕ is the phase of the parent structure factor, and $|F^c_H|$ is the structure amplitude of the heavy atoms alone. The denominator combines the experimental and lack-of-isomorphism error. The latter was the larger component. Derivative data were extended to 2.6 Å, but useful phasing power fell sharply beyond 3 Å.

Diego supplemented the synchrotron parent data at lower resolution (to 3.2 Å). Cooling the crystals to as low as -15 °C provided no significant improvement in lifetime or resolution. A high quality data set was obtained for the parent structure to a resolution of 2.4 Å, although many films produced accurate data to beyond 2.2 Å. There was a modest anisotropy in the intensity distribution that reflected a strong underlying molecular transform in the direction of the DNA helical axis. Table 1 summarizes the data.

The structure was solved mainly by molecular replacement using a modified version of Fitzgerald's program package, MERLOT⁶, applied to the parent data between 8.0-Å and 3.6-Å resolution. The most effective search structure was a variant of the *trp* repressor crystal structure, in which the potential for error in the search structure was diminished. The fact that the two crystal forms of the free *trp* repressor differ significantly in the conformation of their DNA-binding domains⁷, as well as many solvent-exposed side chains, suggested that in the complex these elements could deviate from both crystal structures. The structure factors of the search structure were, therefore, calculated from a composite model where the atoms were placed midway between those of the trigonal and orthorhombic repressor-crystal forms. An atom's contribution to the calculated structure factor was weighted down in proportion to the difference ϵ_i between the atom's position in the two crystal structures. This was done in a resolution-dependent manner by applying to the i^{th} atomic scattering factor a pseudo-thermal term $\exp(-8\pi\epsilon_i^2 \sin^2 \theta / \lambda^2)$. Second, and more important, by subtracting the origin peak from the observed Patterson, we eliminated the effect of the origin term in the calculated component of the translation function. The origin term of the calculated Patterson and its contribution to the translation function increases linearly with the degree of molecular overlap. As correctly placed molecules will never overlap, and incorrectly placed molecules will usually overlap, failure to exclude the origin of the Patterson can significantly raise the noise level of the translation function and obscure the correct solution. With these modifications, the two solutions to the rotation function (first and fifth highest peaks) were 7.9 and 3.9 times the root-

mean-square (r.m.s.) value of the map. The solutions to the translation function were 9.8, 8.1 and 9.6 (cross-translation) times the r.m.s. value. These were the highest peaks in their respective maps and were self-consistent.

A difference map comparing a PbCl₂ derivative and the parent, phased with the molecular replacement solution, showed heavy-atom sites consistent with the non-crystallographic symmetry elements. Phases calculated from the coordinates of the two *trp* repressor molecules, combined with the phases from the PbCl₂ derivative, produced a Fourier map that clearly showed an operator interfaced to each repressor. This initial map (3.0 Å, the limit of the isomorphous replacement phasing was 3.3 Å) was used to build the operator structures and to make the initial adjustments to the repressor model using the computer-graphics program FRODO⁸.

The transformation relating the two independent repressor/operator complexes in the asymmetric unit was established by superimposing the two *trp* repressor molecules found by molecular replacement. We constrained this procedure to superimpose both repressor dyads. This provided the symmetry operations needed to average the electron density corresponding to the four crystallographically independent protomers, namely the four half operator/half repressors. After each stage of refinement the symmetry operators were reassessed with the same procedure using the backbone coordinates of the conformationally invariant 'central core' (residues 18–60)⁷. Remodelling proceeded in two modes; first the model was fitted to the unambiguous portions of the fourfold average electron density. The model was then adjusted to fit the unaveraged density so that local variations would be preserved. In the latter stages of the refinement, the model was fitted mainly to the unaveraged difference maps.

The protein and nucleic acid components of the model were simultaneously refined with a fast Fourier version of a restrained least-squares procedure that combined the programs of Hendrickson and Konnert⁹, Agarwal¹⁰, Finzel¹¹ and Westhof¹². Bond lengths and inter-bond angles were restrained to be, on average, within one standard deviation of their literature values. There were no stereochemical restraints placed on the protein/DNA

interactions, on the relative orientation of the bases, or on any hydrogen bond in the entire complex. Table 1 summarizes the refinement.

Local symmetry implies validity

As the crystallographic asymmetric unit contains two potentially dyad-symmetrical repressor/operator complexes, the crystal structure presents four independent images of the half-repressor/half-operator complex. Except for the amino-terminal arms (residues 2 through 14) and the carboxyl-terminal three residues (106–108), these four protomers are superimposable with an r.m.s. deviation of backbone atoms from their average positions of 0.4 Å. The same degree of similarity applies to the operator structures, as well as to most of the side chains. Unless otherwise stated, the results presented are based on a model that represents the average of the four independent half-repressor/half-operator protomers. To the extent that the four protomers have the same conformation, the structures are free of fortuitous distortions imposed by crystal-packing and are therefore functionally valid representations defined by the intrinsic properties of the complex.

Structure of the repressor

When the repressor structures from the trigonal or orthorhombic crystal forms are superimposed on the repressor structure in the crystalline complex, the conformation in the central core is well conserved. However, the flexible reading heads of the complex showed significant deviations from both crystal forms of the repressor (Fig. 1). This was not unexpected because Lawson *et al.*⁷ have pointed out that the reading heads of the *trp* repressor are flexible, that is, activation of aporepressor to repressor by the binding of L-tryptophan does not impart a rigidly unique structure to the DNA-binding domains of the repressor.

The side chains on the surface of the free repressors are typically mobile. In the repressor/operator complex, however, the temperature factors for the atoms in the side chains that contact the DNA are reduced to the values of the main chain or even lower. Thus, the mechanics of complex formation appear to include the meshing of flexible backbone elements and side chains to form a rigid complementary surface.

The N-terminal 'arms' of the repressor (residues 2 to 14) are poorly ordered and, to the extent that they can be modelled and refined, the four representations in the asymmetric unit assume unrelated conformations defined by interactions with neighbouring protein and DNA molecules and not with their cognate operators. This ill-defined conformation of the amino-terminal residues agrees with the observation made in the crystal structures of the *trp* aporepressor³, pseudorepressor¹³, and two different crystal forms of the repressor^{2,7}, that the conformations of the N-terminal 11 amino acids were largely determined by

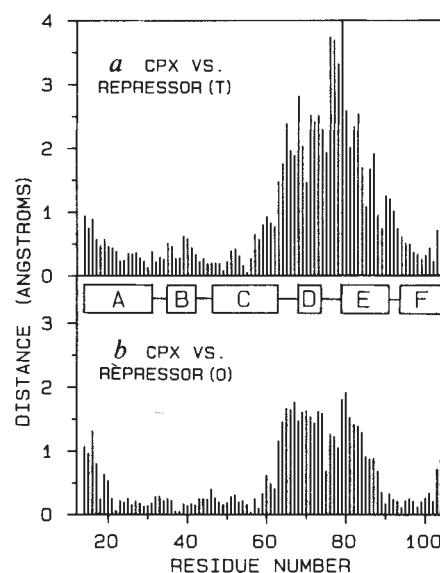


Fig. 1 The *trp* repressor in the complex compared to uncomplexed repressors. The model of the *trp* repressor in the complex was superimposed on models of the *trp* repressor in the trigonal crystal form² (top panel) and in the orthorhombic crystal form⁷ (bottom panel) so as to minimize the r.m.s. deviation of the corresponding α -carbon atoms in the 'core' of *trp* repressor dimer. The ordinate is the deviation of corresponding α -carbon atoms after superimposition. The boxes A through F represent α -helical segments. Helices D and E comprise the helix-turn-helix motif (reading head).

fortuitous packing contacts. The behaviour of this amino-terminal segment reflects the fact that residues 2–15 project out from the surface of the repressor's globular fold.

Carey has shown that chymotryptic removal of amino-acid residues 2 to 7 of both subunits reduces affinity for both operator and nonspecific DNA by between 30- and 100-fold (private communication). A role in DNA binding has been attributed to the presumably flexible N-terminal and C-terminal arms of *cI* and *cro* repressors respectively^{14,15}. The failure of the N-terminal segment to form part of the repressor/operator interface in the crystal may reflect an important difference between the natural interactions and those in the crystal. In the crystal, unlike the operator in its natural setting, DNA fragments are densely packed around the cognate operator offering alternative non-specific binding sites that can be easily reached by the flexible arms. The N-terminal arms might also be involved in protein/protein contacts which stabilize the postulated binding of tandem repressors to overlapping operator sequences in both the *trpEDCBA* and *aroH* operons¹⁶.

Structure of the operator

Table 2 lists the orientation parameters (defined in Fig. 3c) for each base pair or each step between base pairs. The corresponding parameters deduced from the NMR study¹⁷ of a similar symmetrical operator fragment in solution correlate poorly with those of the operator in the complex. The individual deviations of these local parameters from the average are modest in the crystalline complex—indeed, they are smaller than those seen in crystalline uncomplexed DNA. For example, the r.m.s. (and extreme) deviation of the roll and tilt angles for a symmetrical 12-base-pair B-DNA fragment are 5.1° (9.8°) and 1.7° (3.5°) respectively¹⁸. The corresponding values for the *trp* operator fragment in the complex are 4.3° (8.1°) and 2.2° (2.9°). The operator has an average of 10.6 base pairs per turn, with only modest deviations in twist angle (Table 2). The minor groove

Table 2 DNA orientation parameters

Base	Roll (degrees)	Tilt (degrees)	Slide (Å)	Helical twist (degrees)
G ₋₉ /C ₊₉	3.5	-0.8	0.4	24.2
T ₋₈ /A ₊₈	-1.5	-2.7	0.2	42.8
A ₋₇ /T ₊₇	-2.4	-2.0	1.0	32.1
C ₋₆ /G ₊₆	2.6	2.9	0.8	31.8
T ₋₅ /A ₊₅	10.9	2.3	1.0	31.2
A ₋₄ /T ₊₄	6.0	0.4	0.2	35.8
G ₋₃ /C ₊₃	-1.4	2.5	0.9	35.3
T ₋₂ /A ₊₂	1.6	2.9	-0.3	38.7
T ₋₁ /A ₊₁	8.8	0.0	-0.8	31.2
A ₊₁ /T ₋₁				

All parameters were calculated from the fourfold averaged operator with the programs BROLL and CYLIN (R. E. Dickerson, personal communication). The movements are defined in Fig. 3c.

is wider in the *trp* operator (5.5 to 8.1 Å) than in the crystalline dodecamer (3.1 to 7.2 Å) (ref. 18). Its widest point is opposite the parts of the major groove that interact with the reading heads of the repressor. There is also a slight widening of the minor groove about the dyad axis. The range of propeller twists (1.2° to 15.3°) is too small to produce non-Watson-Crick hydrogen bonds between strands. The small local deviations of the operator suggest that there are no focal points for forces distorting the operator, and that the strains of any distortion are distributed.

Figure 3b shows a polar projection of the unit vectors normal to the base pairs. The convex bend in the operator (Fig. 2a) is composed of three straight segments indicated by the three clusters in Fig. 3b. The three segments arise from two symmetrically disposed bends centred eight base pairs apart between position -5 and -4 and between position 4 and 5. This separation is less than the helical periodicity of the DNA; therefore, the molecule will bend in two planes. As a consequence, the DNA appears to be slightly S-shaped when viewed down the molecular dyad. When viewed along the perpendicular to this dyad as in Fig. 2a, the DNA appears C-shaped and follows the protein surface.

The repressor (as seen in the complex) was docked to helically uniform B-DNA. The canonical DNA of this modelled complex was used as a reference point to characterize the deviation of the operator in the complex (Fig. 2b). The atoms comprising bases -5 to +5 superimpose well, with an r.m.s. deviation of 1.4 Å. The phosphate oxygens associated with these bases and contacting the protein superimpose less well, with an r.m.s. deviation of 3.5 Å. In contrast, nucleotides -6 to -9 and +6 to +9 show substantially larger differences in which the phosphates that contact the protein, -7, -8 and -9, deviate by 6.7, 9.1 and 9.6 Å respectively.

The large deviations of the operator from uniform B-DNA are necessary to obtain the 24 phosphate-protein contacts seen in the complex (Fig. 2a). These shifts are due to the cumulative effect of modest local distortions. The largest values of slide and roll occur in the base-pair steps of the sequence ACTAG (from -7 to -3 and its symmetry equivalent), a region that is also the most sensitive to mutations (Fig. 3a). These parameters explain the salient structural features of the DNA. The positive roll in steps T-A and A-G bends the DNA. The steps A-C, C-T and T-A each has a slide of nearly 1 Å, and this run of high slide and roll causes the large displacement of the last three base pairs at both ends of the operator noted above. Thus the segment of operator that carries the strongest sequence determinants for repressor affinity is the segment that exhibits the largest deviations in local orientational parameters and produces the most distinct conformational variation in the repressor/operator complex.

Stabilizing interactions

Hydrogen bonds. There are 14 direct hydrogen bonds between the half-repressor/half-operator (Figs 2a, 4 and 5). All but two involve the unesterified oxygens of six phosphate groups. The exceptions are the two hydrogen bonds in the interaction of Arg 69 with G₋₉, which represent the only direct hydrogen bonds to the functional groups of a base. But these hydrogen bonds are relatively unimportant in specifying the target DNA sequence, for symmetrical mutations of GC₊₉/CG₋₉ have a negligible effect on repressor binding *in vivo*⁴. Four of the protomer's 14 direct contacts to the half-operator are made by amino-acid residues that are not part of the helix-turn-helix motif. The absence of direct hydrogen bonds to mutationally sensitive base pairs implies that, in the *trp* system, specificity is not due to direct hydrogen-bonded contact of the major groove's polar groups by the hydrogen bonding groups of the protein¹⁹.

Six additional solvent-mediated contacts are evident, three to the phosphates and three to the base pairs. One such contact probably involves a bridging calcium ion between the side chain

of Asp 46 and phosphate 1. Two water molecules are positioned to form hydrogen bonds with the repressor and the two unesterified oxygens of phosphate 3; one involves O_γ of Thr 53 and the other O_γ of Thr 81. Phosphate 3 is thereby the focus of four hydrogen-bonded interactions, two mediated by the solvent and at least two directly from the protein (Arg 54 and the ring nitrogen of the bound tryptophan).

There are three well-ordered water molecules that mediate hydrogen-bonded interactions between the half-repressor and the base pairs of the half-operator. Figure 5c-e shows how two of the water molecules bridge the peptide nitrogens of Ile 79 and Ala 80 at the amino terminus of the E-helix to the hydrogen-bond acceptors of the purines A₅ and G₆. The water that bridges N79 to N7 of G₆ appears to be firmly positioned by an additional interaction with the N_ε of Lys 72. This arrangement of hydrogen bonds channels the positive potential of the E-helical dipole to the bases that face the tip of the helix-turn-helix. The third water molecule links the γ-hydroxyl of Thr 83 to N6 and N7 of A₇. Figure 3 shows that the bases involved in these water-mediated interactions are among the most important in specifying the repressor's affinity for the operator sequence.

Non-polar contacts. The α-carbon of Gly 78, the fourth residue of the turn in the helix-turn-helix, is close enough to the hydrophobic surface of the DNA (C6 of T₄ and C8 of A₅; Fig. 5e) to exclude water; however, this contact has no discriminatory function because other bases can be accommodated without steric clashes. Furthermore, the region of the operator that opposes the α-C of Gly 78 is hydrophobic, irrespective of the DNA sequence. The contact is close enough, however, to explain why amino acids other than glycine should disrupt repressor binding²⁰. The C_γ of Thr 83 forms a non-polar contact to the C2' of A₋₇. Interestingly, the non-polar side chains of Ile 79 and Ala 80, positions postulated to be important in DNA recognition²¹, do not make extensive contact with non-polar surfaces of the operator. Instead, they face mainly the hydrogen-bonded functional groups of base pairs ±6 and ±7 (Fig. 5c). The C_{γ1} (methylene) of Ile 79 does contact the C5 methyl of T₈, but apparently does not serve a discriminatory function because this position is not conserved in the three operators of the *trp* system.

The role of bound L-tryptophan. By comparing the high-resolution crystal structures of the unliganded aporepressor with that of the liganded active repressor, Zhang *et al.*³ have shown that the main effect of the corepressor, L-tryptophan, is to orient the flexible reading heads so that the helix-turn-helix motif can penetrate successive major grooves of the operator. Figures 2a and 5f and g show that, in addition to the 'allosteric' effect on the protein's conformation, bound L-tryptophan and the neighbouring residues of the pocket interact directly with DNA. The indole ring nitrogen of L-tryptophan forms a hydrogen bond to phosphate 3. The 'floor' of the indole ring's binding pocket is formed by the proximal methylene groups of Arg 54 whose guanidino group makes a hydrogen bond (possibly two) with phosphate 3. L-tryptophan's α-carboxylate makes a bifurcated hydrogen bond which positions the guanidino group of Arg 84 so that it interacts with phosphate 2. The corepressor's role in forming these contacts depends on an unusual binding mode, in which tryptophan's aliphatic side chain is maintained under torsional strain by strong contacts with the protein^{13,22}. These contacts, combined with the direct interactions of the ligand and its binding site with the operator, cause the bound tryptophan to have the lowest thermal parameters in the entire structure; thus, the corepressor is the most firmly fixed element in the complex.

Solvent excluded interface. The flexibility of both the DNA-binding domains and the operator permit a moulding of the contact surface that makes 2,900 Å² inaccessible to solvent at the interface of the *trp* repressor/operator complex. This large solvent-excluded surface is typical of subunit interfaces in hetero-oligomeric proteins where structural adjustments are made on aggregation²³. Nearly all the polar groups that become

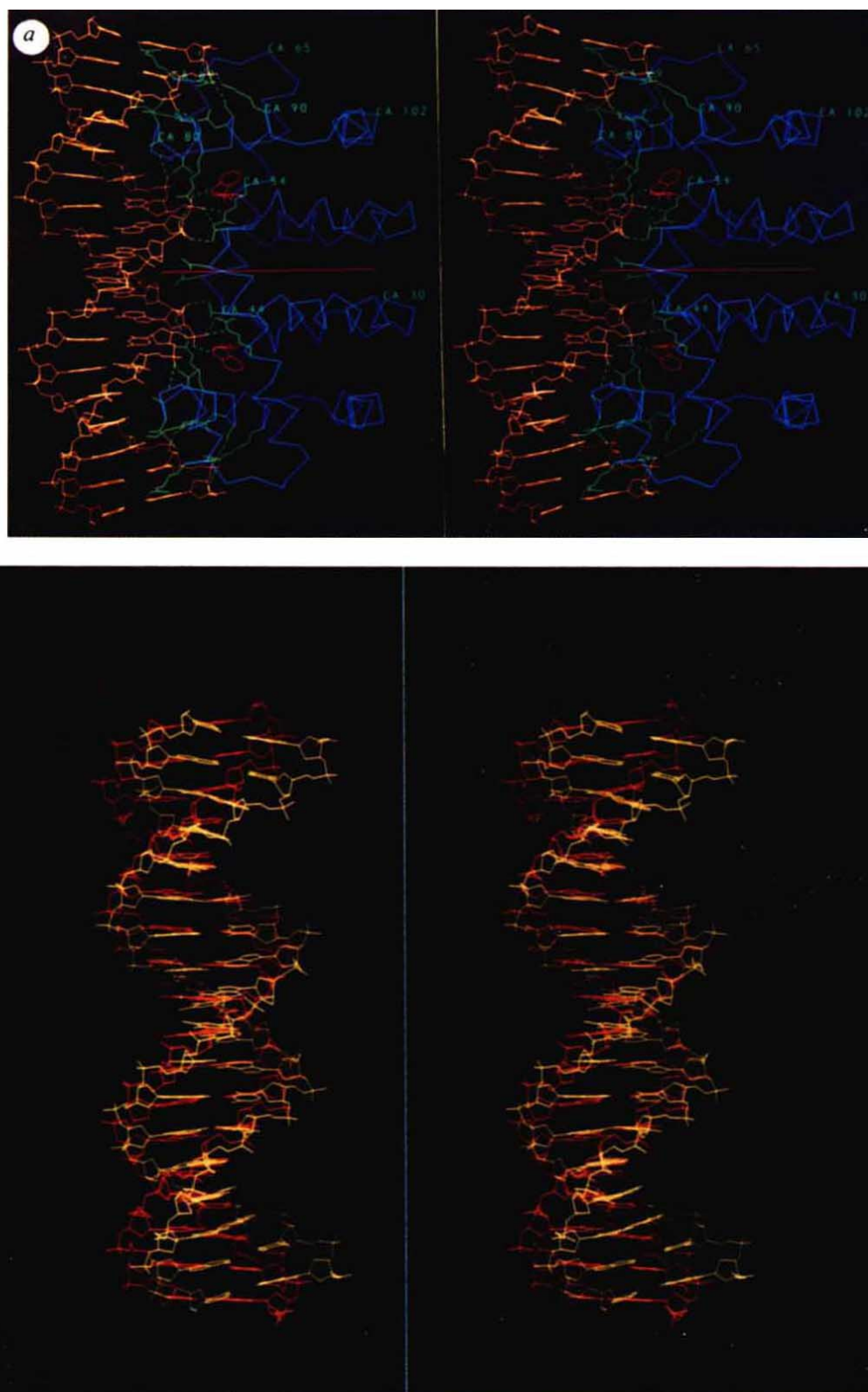


Fig. 2 The *trp* repressor/operator complex. *a*, The crystal structure. An α -carbon trace of the repressor backbone (blue) plus the side chains (green) that make direct hydrogen bonds (dashed line) to the operator (gold). *b*, The DNA structure. Helically uniform B-DNA (orange) is contrasted to the DNA as seen in the complex (gold).

solvent-inaccessible are hydrogen-bonded and no phosphate group is completely buried by protein. Assembly of protein subunits is favoured by the entropic effect of mobilizing water from the buried surfaces, which more than compensates for the restrictions imposed by stable tertiary and quaternary structures²⁴. Record²⁵ has noted that mobilization of water is also one of the main contributors to the stabilization of specific protein/nucleic-acid complexes. The release of water would be expected to contribute a large entropic component to the stabiliz-

ing free energy of the *trp* repressor/operator interaction in view of the substantial solvent-excluded surface seen in the crystalline complex. The stereochemical adjustment of both the protein and nucleic-acid components of the complex appear designed to exploit this mode of stabilization.

***trpR* mutations.** Changes in amino-acid side chains can diminish operator affinity in three ways; by depriving the complex of an important stabilizing interaction, by introducing a side chain that sterically or electrostatically disturbs the contact surface,

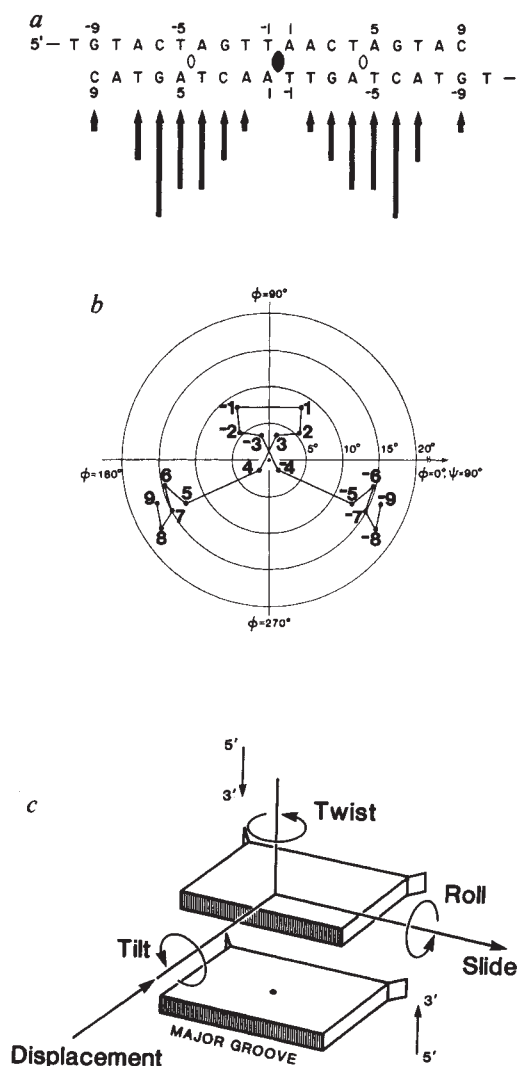


Fig. 3 Base-pair orientations. *a*, The sequence of the duplex DNA fragment in the co-crystals that simulates the operator. Symmetry was achieved by changing the *trpEDCBA* operator sequence from (A·T)₋₈ to (T·A)₋₈ to match (A·T)₊₈. This does not affect repressor binding⁴. Numbering of base pairs is from 5' to 3' on the top strand from an origin at the twofold rotational axis of symmetry, so the dyad axis is between -1 and +1. The numbering of nucleotides on the bottom strand retains dyad symmetry. The arrows indicate the nucleotides where mutations diminish repressor affinity. The size of the arrow is a rough measure of the base pair's importance in specifying repressor affinity, as deduced from the preservation of the nucleotides in the three operator sequences of *Escherichia coli* (*trpEDCBA*, *trpR* and *aroH*) and the directed mutational studies of Bass *et al.*⁴ For example, any symmetrical substitution for (G·C)₆ or (C·G)₋₆ totally disrupts operator binding in the assay used by Bass *et al.* *b*, A polar projection (on to the equatorial plane) of unit vectors normal to each base pair's average plane. Base pairs are designated by the number of the base in the top strand of *a*. Coordinates are the latitude (polar angle ψ) measured from the pole, and longitude ϕ measured counter-clockwise from the complex's dyad (which has the coordinates $\phi = 0^\circ$, $\psi = 90^\circ$). The pole represents the overall axis of the DNA. A straight segment of B-DNA produces a cluster around the segment's axis. An abrupt bend forms a separate new cluster. *c*, Definition of relative base-pair orientational angles and 'slide' used in Table 2. The arrows indicate the positive sense. The angles describe the change in orientations proceeding from one base pair to the next (a 'step'). Slide and displacement refer to a movement in relation to the 'local' helical axis defined by a base pair and its immediate neighbour. The orthogonal coordinate basis is defined by a local normal to the average plane of the two adjacent base pairs of the step, and the bisector of the base pair to be rotated.

or by disrupting the structure of the repressor. Examination of the crystalline complex enables us to explain all the negative missense *trpR* mutations²⁰ in terms of one or more of these effects. Deleterious mutational changes have been shown to occur at most of the positions in contact with the operator. Figure 4c shows the correspondence between amino-acid sequence positions in the helix-turn-helix, where deleterious missense mutational changes occur, and the amino-acid side chains that contact the operator.

Electrostatic attraction. Warwicker (personal communication) calculated the electrostatic charge potential surrounding the repressor. The positive charge potential distribution clearly follows the contours of B-DNA, suggesting that electrostatic field effects may contribute significantly to interaction with the operator. This view is supported by the observation that mutations such as Glu 18 to Lys increase repression²⁰, even though the mutationally inserted lysine side chains cannot possibly contact the operator. We suggest that the tight fit of the specific complex strengthens this general electrostatic attraction by diminishing the electrostatic screening effect of the solvent.

Tandem binding. Kumamoto *et al.*¹⁶ proposed, on the basis of methylation-protection and footprinting experiments, that one *trp* repressor dimer binds the *trpR* operator and that more than one repressor might bind to overlapping operator-like sequences in both the *trpEDCBA* and *aroH* operons. The relationship of the tandemly overlapping repressors can be best described as a strict twofold rotation of the entire repressor around an axis placed between base pairs 4 and 5, as shown by the open symbols in Fig. 3a. A 180° rotation of the entire repressor around this axis (with no further adjustments) places the 'second' repressor in a position that preserves the protein-phosphate contacts described above. In this model, no stereochemical clashes occur that could prevent this tandem binding.

Discussion

Different B-DNA sequences have characteristic distributions of hydrogen-bond donor and acceptor groups on the bases, but rather similar sugar-phosphate backbones. This fact has led to a postulate that specific affinity of proteins for their DNA targets arises from complementary protein-DNA hydrogen bonding through groups in the major and minor grooves of the nucleic acid¹⁹. This 'direct-readout' mechanism for recognition can be extended to include polar contacts mediated through ordered water molecules.

A second principle, called 'indirect readout', relies on the increasingly recognized fact that the B-DNA exhibits a large degree of sequence-dependent structural variation²⁶⁻²⁹. Arguments stemming from this observation have been used to explain the positioning of DNA in nucleosomes^{30,31,33}, the specificity of nucleases^{33,34}, the bending of DNA³⁵⁻⁴¹ and the preferential binding of *lac* repressor to poly(dA·dT) (ref. 42). In the indirect readout mechanism, the primary function of the DNA sequence is to permit a conformation that substantially enhances attractive interaction with its cognate protein. Direct and indirect readouts do not invoke mutually exclusive roles for the base sequence, and clearly both principles can confer specificity to various degrees in any particular complex.

Two unrefined crystal structures of specific protein/DNA complexes have been reported at resolutions that provide some insight into the stereochemistry of specific protein/DNA interactions. Rosenberg and his colleagues pointed out that significant protein-induced DNA-deformation, as well as potential hydrogen-bonding interactions with the functional groups of the major groove, are important in forming the complex between the *EcoRI* restriction endonuclease and its six-base-pair substrate⁴³. Harrison and his colleagues suggested that in the crystalline complex of *p434* repressor and its 14-base-pair operator, the non-contacted bases at the centre of the operator may govern the degree of helical twist needed to accommodate both polar and

Fig. 4 Residues involved in direct contacts between *trp* repressor and operator. *a*, Schematic view of the upper half of *trp* operator as 'viewed' by the *trp* repressor showing operator's phosphates (circles with numbers) and the repressor's functional groups (arrows) that form direct hydrogen bonds. *b*, Schematic view of the amino-acid positions (circled amino-acid numbers) in the helix-turn-helix that make direct hydrogen bonds to the operator (shown by arrows) as seen in Figs 1 and 4*a*. The figure emphasizes that the second helix, or E-helix, of the bihelical motif is 'pointed into' the major groove with its α -helical axis almost perpendicular to the global DNA axis. The E-helix does not 'lie in' the major groove with its axis parallel to the groove. The amino-acid sequence of the helix-turn-helix. The circled residues make direct contact to the operator. The dots indicate residues which when altered by mutation significantly diminish repression²⁰. The underlined residues are involved in stabilizing the conformation of the DNA-binding domain. The sequences of other bacterial regulatory proteins are shown for comparison²⁷.

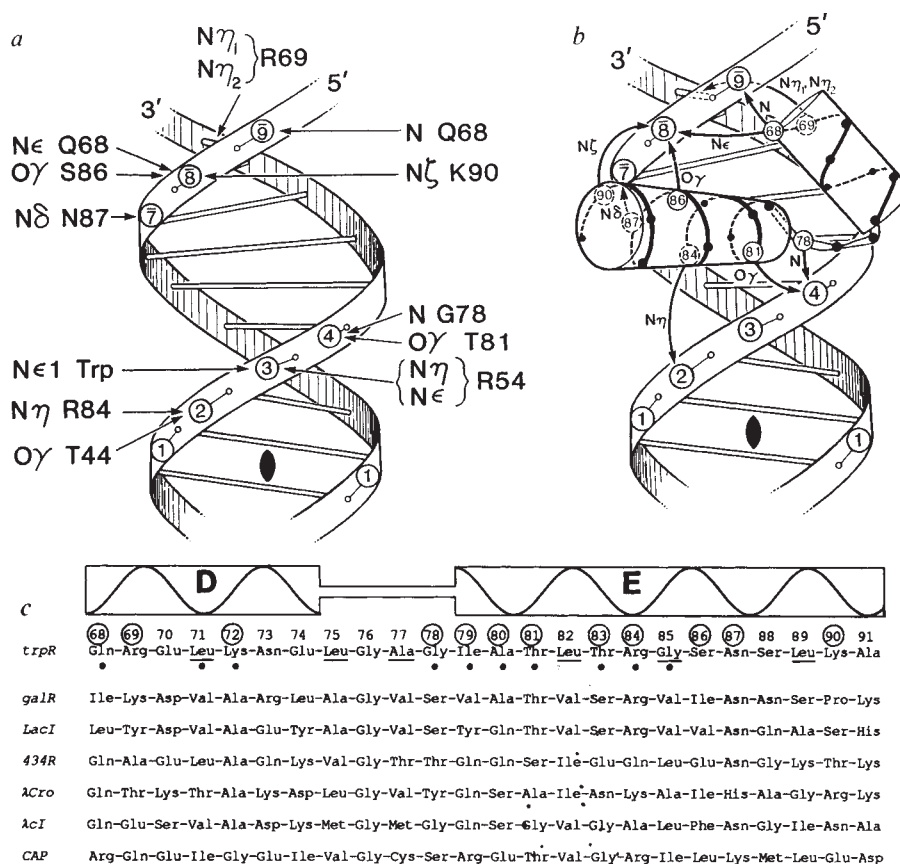
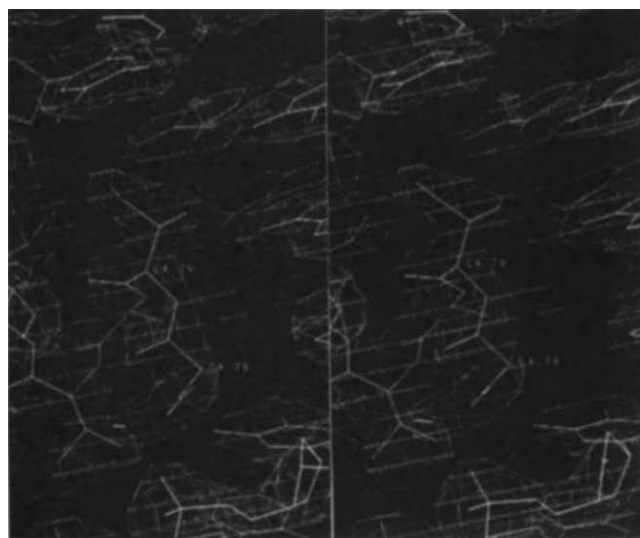


Fig. 6 Electron density map. A Fourier synthesis containing coefficients $[m(r|F_0| - (r-1)|F_c|)\exp(\phi)]_h$, where ϕ is the centroid phase and m the figure-of-merit of a joint-phase probability distribution with contributions from both the refined model and the PbCl_2 derivative. The scaling term r is ~ 2.5 and adjusts the relative contributions to minimize model bias in the maps. The map is averaged fourfold as described in the text. The model shows the amino-terminal end of the E-helix deep in the operator's major groove. Note the density for a water molecule near N80.



non-polar stabilizing contacts^{44,45}. The graded affinity for tandem operators is at the heart of the biological role of the repressors in this family of temperate phages, and it is mainly the central base pairs that distinguish the operator sequences.

In this study of the *trp* repressor/operator complex, we have sufficient resolution, accuracy and redundancy to state that there are no direct hydrogen bonds between the functional groups of the bases and the protein that can account for the specific affinity of the *trp* repressor for the operator sequence. The only direct hydrogen bonds within the major groove are between Arg 69 and G₋₅; however, for reasons pointed out earlier, this interaction contributes little to the specificity. Similarly, there are also no van der Waals' contacts between the non-polar side chains

of the protein and the non-polar surfaces of the bases that can account for specificity. In fact, the contact surface does not appear to be able to exclude non-operator DNA sequences by steric interference, such as that produced by a snug contact with a C5 of cytosine that would clash with a 5-methyl group were it replaced by a T. It should be emphasized that our failure to observe direct protein contacts with the contours of the major groove does not arise from our inability to see the side chains in the electron density. With the exception of the N-terminal arms, all the amino acids that could conceivably contact the DNA bases have refined robustly with a high degree of positional certainty and local symmetry. Figure 6 illustrates the typically strong electron density for the side chains of the contact surface.

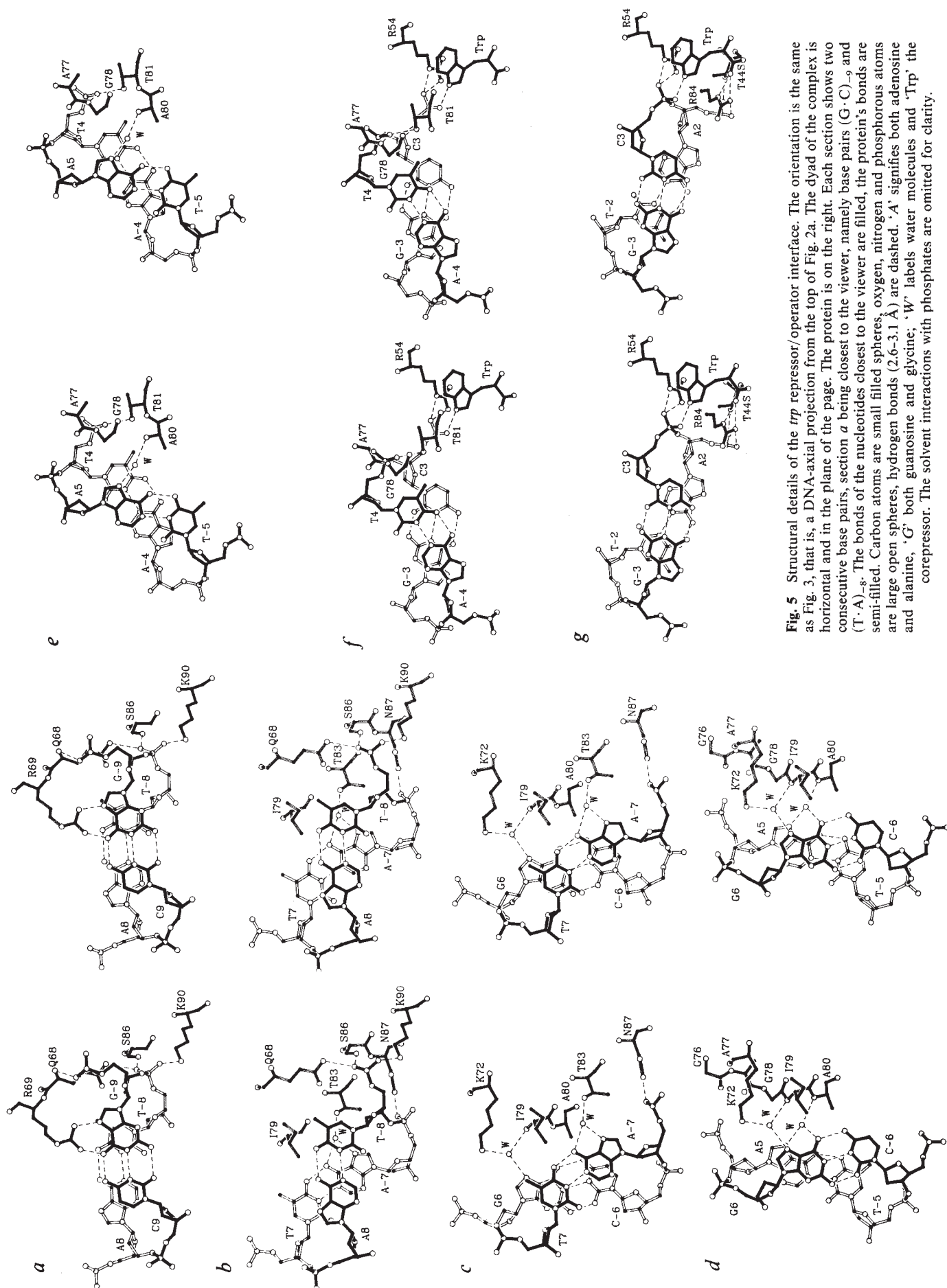


Fig. 5 Structural details of the *trp* repressor/operator interface. The orientation is the same as Fig. 3, that is, a DNA-axial projection from the top of Fig. 2a. The dyad of the complex is horizontal and in the plane of the page. The protein is on the right. Each section shows two consecutive base pairs, section *a* being closest to the viewer, namely base pairs (G·C)₋₉ and (T·A)₋₈. The bonds of the nucleotides closest to the viewer are filled, the protein's bonds are semi-filled. Carbon atoms are small filled spheres, oxygen, nitrogen and phosphorus atoms are large open spheres, hydrogen bonds (2.6–3.1 Å) are dashed. 'A' signifies both adenine and alanine, 'G' both guanine and glycine; 'W' labels water molecules and 'Trp' the corepressor. The solvent interactions with phosphates are omitted for clarity.

We have, however, identified three polar contacts mediated by well ordered water molecules that bridge the functional groups of mutationally sensitive base pairs⁴ to the protein and thereby contribute to 'direct readout' of the operator sequence (Fig. 5b-d). Two of the water molecules (contacting N79 and N80) are compatible only with a G at position 6 and a purine at position 5. These restrictions are completely consistent with the result of a directed mutational analysis⁴, in which symmetrical substitution of any base for G at position 6 or of a pyrimidine at position 5 maximally reduced *in vivo* operator affinity. The pattern of mutational effects at position 7 can be similarly rationalized by the solvent-mediated contacts between A₇ and Thr 83. Moreover, repression is diminished by *trpR* mutations that change the residues whose side chains are involved in these water-mediated interactions, Lys 72 (S. Bass, V. Sorrells and P. Youderian, private communication) and Thr 83 (ref. 20). It should be noted that water-mediated contacts cannot explain the importance of base pairs 3 and 4. Moreover, there is no assurance that alternative patterns of water-mediated contacts could not serve to bridge non-operator sequences to the repressor. What then accounts for the specificity of the *trp* repressor/operator complex?

We believe that in the *trp* repressor/operator system a primary mode of sequence specific recognition is 'indirect readout'; that is, the operator sequence permits the DNA to assume a characteristic conformation that will make 24 direct hydrogen bonds to the protein through its phosphate groups and increase the solvent-excluded surface. It is obvious from docking experiments that straight uniform DNA with average conformational parameters can support only a small fraction of such a hydrogen-bonded network. The fact that four of the hydrogen-bonded phosphates in the protomer accept more than one hydrogen bond, and that two of these are caged by three direct hydrogen bonds, requires that the DNA backbone should assume a very precise geometry. Those conformational adjustments also enlarge the solvent-excluded contact surface to one that is comparable to those between subunits in stable oligomeric proteins. Moreover, exclusion of the solvent's screening effect is likely to

augment both general and local electrostatic attractions.

We propose that DNA sequences other than an operator sequence, such as those containing operator constitutive mutations, might achieve this favourable conformation, but at too high a cost in internal energy to form a stable complex. *In vitro* measurements of the *trp* repressor specificity⁴⁶ show a 10⁴-fold preference for the operator over a random non-operator site. This implies a difference of at least 5 to 6 kilocalories per mole in the free energy of binding. It is likely that the water-mediated hydrogen bonds involving base pairs 5 through 7 contribute part of the free-energy difference that targets *trp* repressor to the *trp* operator. Could the bulk of the free-energy difference responsible for operator-specificity be accounted for by sequence-dependent differences in the work required to achieve the conformation of the operator in the complex? Substantial differences in stacking energies have been calculated for small changes in the orientation of adjacent base pairs in B-DNA^{28,47,48}. These variations in energy are highly dependent on the species of neighbouring base pairs, and might easily account for strong sequence-specific restraints on the required adjustments to the DNA's conformation.

We thank P. M. D. Fitzgerald for her program MERLOT; R. L. Fox, W. Mandecki, M. Hayden, R. Metzger and B. Powell (Abbott Laboratories) for synthesizing the oligonucleotides; P. Phizakerley and E. Merritt (Stanford Synchrotron Radiation Laboratory); W. Schildkamp and K. Moffat (Cornell High Energy Synchrotron Source), and N.-g. Xuong and others at the Area Detector Facility (UCSD) for providing instrumental resources for data collection; J. Carey, P. Youderian, S. Bass and R. Gunsalus for providing genetic and biochemical information before publication; C. Yanofsky, V. Horn and J. Paluh for overproducing strains; H. Drew, M. McCall, C. Calladine, R. Dickerson and W. Olson for help with the analyses of the DNA structure; and finally, G. Sechler, N. Freed, J. Cohen and C. Anderson for help in preparing the manuscript. This work was supported by grants from the USPHS and the ACS. C.L.L. was a predoctoral trainee of the USPHS, and B.F.L. is a Fellow of the Jane Coffin Childs Memorial Fund for Cancer Research.

Received 6 June; accepted 9 August 1988.

- Somerville, R. L. *Amino Acid Biosynthesis and Regulation* (ed. Hermann, R. M. & Somerville, R. L.) 351-378 (Addison-Wesley, Reading, Massachusetts, 1983).
- Schevitz, R. W., Otwinowski, Z., Joachimiak, A., Lawson, C. L. & Sigler, P. B. *Nature* **317**, 782-786 (1985).
- Zhang, R. *et al. Nature* **327**, 591-597 (1987).
- Bass, S., Sugiono, P., Arvidson, D. N., Gunsalus, R. P. & Youderian, P. *Genes and Development* **1**, 565-572 (1987).
- Joachimiak, A. *et al. J. biol. Chem.* **262**, 4917-4921 (1987).
- Fitzgerald, P. M. D. *J. appl. Crystallogr.* **21**, 273-278 (1988).
- Lawson, C. L. *et al. Proteins* **3**, 18-31 (1988).
- Jones, T. A. in *Computational Crystallography* (ed. Sayre, D.) 303-317 (Clarendon, Oxford, 1982).
- Hendrickson, W. A. & Konert, J. H. *Structure, Conformation, Function and Evolution* Vol. 1 (ed. Srinivasan, R.) 43-57 (1981).
- Agarwal, R. C. *Acta Crystallogr. A* **34**, 791-809 (1978).
- Finzel, B. C. *J. appl. Crystallogr.* **20**, 53-57 (1987).
- Westhof, E., Dumas, P. & Moras, D. *J. molec. Biol.* **184**, 119-145 (1985).
- Lawson, C. L. & Sigler, P. B. *Nature* **233**, 869-871 (1988).
- Pabo, C. O., Krouatin, W., Jeffrey, A. & Sauer, R. T. *Nature* **298**, 441-443 (1982).
- Caruthers, M. H. *et al. in Protein Structure Folding and Design* (ed. Oxender, D. L.) 221-228 (Liss, New York, 1986).
- Kumamoto, A. A., Miller, W. G. & Gunsalus, P. *Genes and Development* **1**, 556-564 (1987).
- Lefevre, J. F., Lane, A. N. & Jardetzky, O. *Biochemistry* **26**, 5076-5090 (1987).
- Dickerson, R. E., Kopka, M. L. & Pjura, P. *Biological Macromolecules and Assemblies* Vol. 2 (ed. McPherson, A. & Jurnak, F.) 38-126 (1985).
- von Hippel, P. H. & Berg, O. G. *Proc. natn. Acad. Sci. U.S.A.* **83**, 1608-1612 (1986).
- Kelley, R. L. & Yanofsky, C. *Proc. natn. Acad. Sci. U.S.A.* **82**, 483-487 (1982).
- Ebright, R. H., Cossart, P., Gicquel-Sanzey, B. & Beckwith, J. *Proc. natn. Acad. Sci. U.S.A.* **81**, 7274-7278 (1984).
- Marmorstein, R. Q., Joachimiak, A., Sprinzl, M. & Sigler, P. B. *J. biol. Chem.* **262**, 4922-4927 (1987).
- Chothia, C., Wodak, S. & Janin, J. *Proc. natn. Acad. Sci. U.S.A.* **73**, 3793-3797 (1976).
- Janin, J. & Chothia, C. *FEBS: 12th Meeting, Dresden* **52** (ed. Hoffman, E., Pfeil, W. & Aurich, H.) 227-237 (Pergamon, Oxford, 1978).
- Record, M. T. in *Unusual DNA Structures* (ed. Wells, R. D. & Harvey, S. C.) 237-252 (Springer, New York, 1988).
- Dickerson, R. E. & Drew, H. R. *J. molec. Biol.* **149**, 761-786 (1981).
- Ohlendorf, D. H., Anderson, W. F. & Matthews, B. W. *J. molec. Biol.* **19**, 109-114 (1983).
- Haran, T. E., Berkovich-Yellin, Z. & Shakked, Z. *J. Biomolec. Struct. Dynam.* **2**, 397-412 (1984).
- Calladine, C. R. & Drew, H. R. *J. molec. Biol.* **192**, 907-918 (1986).
- Satchwell, S. C., Drew, H. R. & Travers, A. A. *J. molec. Biol.* **191**, 659-675 (1986).
- Drew, H. R. & Calladine, C. R. *J. molec. Biol.* **195**, 143-173 (1987).
- Travers, A. A. & Klug, A. *Phil. Trans. R. Soc. B* **317**, 537-561 (1987).
- Lomonosoff, G. P., Butler, P. J. G. & Klug, A. *J. molec. Biol.* **149**, 745-760 (1981).
- Suck, D., Lahm, A. & Oefner, C. *Nature* **332**, 464-468 (1988).
- Marini, J. C., Levene, S. D., Crothers, D. M. & Englund, P. T. *Proc. natn. Acad. Sci. U.S.A.* **79**, 7664-7668 (1982).
- Wu, H.-M. & Crothers, D. M. *Nature* **308**, 509-513 (1984).
- Bossi, L. & Smith, D. M. *Cell* **39**, 643-652 (1984).
- Snyder, M., Buchman, A. R. & David, R. W. *Nature* **324**, 87-89 (1986).
- Zahn, K. & Blattner, F. R. *Science* **236**, 416-427 (1987).
- Stenzel, T. T., Patel, P. & Bastia, D. *Cell* **49**, 709-717 (1987).
- Nelson, H. C. M., Finch, J. T., Luisi, B. F. & Klug, A. *Nature* **330**, 221-226 (1987).
- Klug, A. *et al. J. molec. Biol.* **131**, 669-680 (1979).
- Frederick, C. A. *et al. Nature* **309**, 327-331 (1984).
- Anderson, J. E., Ptashne, M. & Harrison, S. C. *Nature* **326**, 846-852 (1987).
- Harrison, S. C. *et al. Bull. Inst. Pasteur* **86**, 55-64 (1988).
- Carey, J. *Proc. natn. Acad. Sci. U.S.A.* **85**, 975-979 (1988).
- Ornstein, R. L., Rein, R., Breen, D. L. & MacElroy, R. D. *Biopolymers* **17**, 2341-2360 (1978).
- Gotoh, O. & Tagashira, Y. *Biopolymers* **20**, 1033-1042 (1981).

Rapid infrared and optical variability in the bright quasar 3C273

Thierry J.-L. Courvoisier*[¶], E. Ian Robson[†],
Andre Blecha[‡], Patrice Bouchet[§], David H. Hughes[†],
Kevin Krisciunas^{||} & Hugo E. Schwarz[§]

* ST-ECF, ESO, Karl Schwarzschild Strasse 2,
D-8046 Garching bei Munich, FRG

[†] School of Physics and Astronomy, Lancashire Polytechnic, Preston,
PR1 2TQ, UK

[‡] Observatoire de Geneve, CH-1290 Sauverny, Switzerland

[§] ESO, Casilla 19001, Santiago 19, Chile

^{||} JAC, 665 Komohana St, Hilo, Hawaii 96720, USA

The magnitude and timescale of variations in the luminosity of quasars provide crucial constraints on the mechanism of emission and the physical size of the emitting region. We have observed variations by a factor of two in the infrared flux from the bright quasar 3C273 on a timescale as short as one day. Variations of this magnitude, over such short periods, are unprecedented. In February 1988, the behaviour of the source changed from having a stable infrared flux and slow optical variations (on timescales of weeks or months) to a state characterized by recurrent infrared and optical flaring. The optical variations were of several per cent per day, changing from increase to decrease approximately every week. The amplitude of the repeated optical flares was 30–40%. The data are consistent with re-injection/acceleration of electrons followed by rapid cooling. The inferred magnetic field is 0.7 gauss and the data are marginally consistent with no relativistic beaming.

The observations described here are part of a continuing multi-frequency study of the relationships between the different emission components of 3C273 (ref. 1). This study has followed, over a long period of time, the emission from 3C273 over a frequency range from radio to X rays, and has already revealed several interesting types of behaviour: a decrease in the 2–10 keV flux by 40% in 20 days in early 1984¹, a drop in the ultraviolet emission by 40% in approximately six months in 1987² and a decrease in the flux for wavelengths > 10 microns by factors of two to four in early 1986³. Robson and co-workers observed an

infrared-to-millimetre flare in 1983⁴, which was subsequently modelled by Marscher and Gear⁵. The behaviour described in this letter differs from all these previous events by the violence of the phenomena involved and by the dense (albeit still insufficient) time- and frequency-sampling of the variations. Previous reports of fast variability in 3C273 include a factor of two variation in the X-ray flux in ~0.5 day (ref. 6) and millimetre variability on timescales of ~1 day (ref. 7).

The optical observations were conducted at the 70-cm Swiss telescope in La Silla. All data points are averages of two measurements 30 min apart. Measurements have an uncertainty of 1–2%, mainly because of counting statistics. The absolute calibration of the Geneva photometry has been revised recently⁸ and this revision is included in the data discussed here. The near-infrared observations were obtained using the 2.2-m and 1-m telescopes at ESO, and UKIRT and IRTF at Mauna Kea. The infrared photometry has been reduced to the zero-magnitude values published in the IRTF photometry manual⁹. The overall calibration factors are claimed to be good to within 10%, but photometric uncertainties are much less, usually ~3–5% at JHK. We obtained only a few 4.8-, 10- and 20-micron data points, which were converted into flux density values as for the near-infrared measurements. Sub-millimetre and millimetre data were also collected during this study from UKIRT, SEST and JCMT.

The sampling of the light curve is best in the optical, with at times daily observations over the period of activity. Gaps in the data are caused by the bright Moon being too close to the object and by periods of bad weather. In the infrared the sampling frequency is less, although daily observations were available in JHK around the important event of 9 March. Even with this very regular temporal and spectral coverage, the density of the sampling remains the limiting factor in the description of the variations.

Between February and April 1988 five maxima were observed in the *V* flux of 3C273 (Fig. 1a), separated on average by 15 days. Two of these maxima, however, are separated by only three days (10 and 13 March). Compared with the flux at the beginning of the year (29 mJy in *V*, 26 mJy in *B*, 25 mJy in *U*), the peak fluxes measured on 10 March were 36.5 mJy (26% above the January flux) in *V*, 31.9 mJy (23% above the January flux) in *B* and 31.1 mJy (19% above January) in *U*. The maximum on March has a lower amplitude in *B* and *U* (see Table 1).

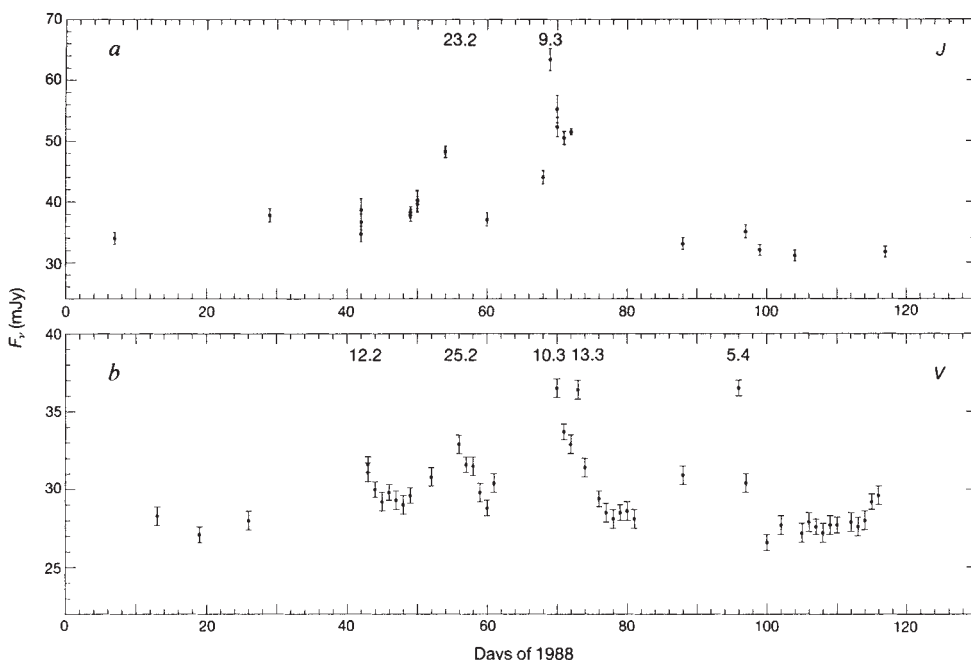


Fig. 1 Optical (a) and infrared (b) light curves of 3C273 from January to mid-April 1988 showing large variability on timescales as short as a day.

[¶] Present address: Observatoire de Geneve, CH-1290 Sauverny, Switzerland.

Table 1 Fluxes (mJy) and slopes observed during the 9 March flare

Date	UT*	U	B	V	J	H	K	alpha (UBV)†	alpha (JHK)†
8.3	5 h				45.6	68.3	111.0		1.40
					1.0	1.0	2.0		
9.3	6 h				65.8	87.6	137.0		0.92
					2.0	2.0	2.0		
10.3	6 h	31.1	31.9	36.5	57.4	83.2	129.0	0.66	1.15
	6 h				2.0	2.0	2.0		
10.3	6 h				54.3	82.3	130.0		1.42
					2.0	2.0	2.0		
11.3	14 h	28.7	30.1	33.7	48.0	68.2	112.8	0.78	1.20
	6 h				1.4	2.0	3.4		
12.3	13 h	27.5	28.1	32.9	48.9	71.1	116.0	1.15	1.26
	6 h				1.5	2.1	3.5		
13.3		28.7	30.0	36.4				1.50	
	5 h								
14.3		26.9	27.3	31.4				0.95	
	6 h								

* The two UTs refer to infrared measurement (top) and optical measurement (bottom).

† The slopes refer to the flare component after subtraction of the quiescent emission.

The optical maximum observed on 5 April was characterized by a different spectral distribution, with a larger flare amplitude in *U* and *B* than in *V*. The fluxes were then 36.5 mJy (26% above the January flux) in *V*, 36.4 mJy (40% above January) in *B* and 33.4 mJy (34% above January) in *U*. The fastest variations were measured as the spectrum fell but insufficient observations before the flare restrict the limits we can impose on the rate of rise of flux. The maximum night-to-night decreases were 5 mJy in *V*, 8 mJy in *B* and 6 mJy in *U*. This corresponds to a rate of change in the luminosity of $\sim 6 \times 10^{40} \text{ erg s}^{-2}$ (assuming isotropic emission and $H_0 = 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$) over a bandwidth of 100%.

Two maxima can be seen in the infrared data on 23 February and 9 March (Fig. 1b), the *J* flux rising to 46 mJy and doubling to 65 mJy respectively. This latter event was extremely rapid: the infrared flux increases were 20 mJy (*J*), 19 mJy (*H*) and 26 mJy (*K*) during 25 h from 8 to 9 March. The *J* flux change corresponds to $\sim 6 \times 10^{40} \text{ erg s}^{-2}$ over a 100% bandwidth, a similar absolute value to that measured in the optical. The observed fluxes are given in Table 1.

Preliminary analysis of the sub-millimetre and millimetre data shows a substantial increase of the flux in February and March over the values observed earlier in the year. Measurement of the rate of change in this spectral region will have to await an absolute flux calibration for the different instruments used and will be presented at a later date.

Normally 3C273 shows no evidence to support any intrinsic optical polarization; typical values are $0.3 \pm 0.13\%$ (*U*), $< 0.17\%$ (*B*), $0.24 \pm 0.8\%$ (*V*), $0.34 \pm 0.08\%$ (*R*) and $0.39 \pm 0.09\%$ (*I*) as measured on 27 March 1987 (J. Rodriguez, private communication). We obtained daily optical polarization data for seven days from 24 February (Fig. 2). The polarization varies from $0.7 \pm 0.2\%$ to $2.5 \pm 0.2\%$ in *R* and *I*, and from $0.8 \pm 0.2\%$ to $1.8 \pm 0.2\%$ in *U*, *B* and *V*. Both amplitude and position angle of the polarization vary from night to night (Fig. 2). A *K*-polarization measurement on 3 March indicated that the level of polarization had decreased to normal values between the two infrared flares (T. Jones, private communication).

Because of the very fast time variability, great care has to be taken in attempting to construct representative optical-infrared spectra of the source. Simultaneous optical and infrared data are available for 10 March (optical at 5:55 UT, infrared at 6:00 UT), the day following the maximum observed infrared intensity. These data are shown in Fig. 3 together with a set of

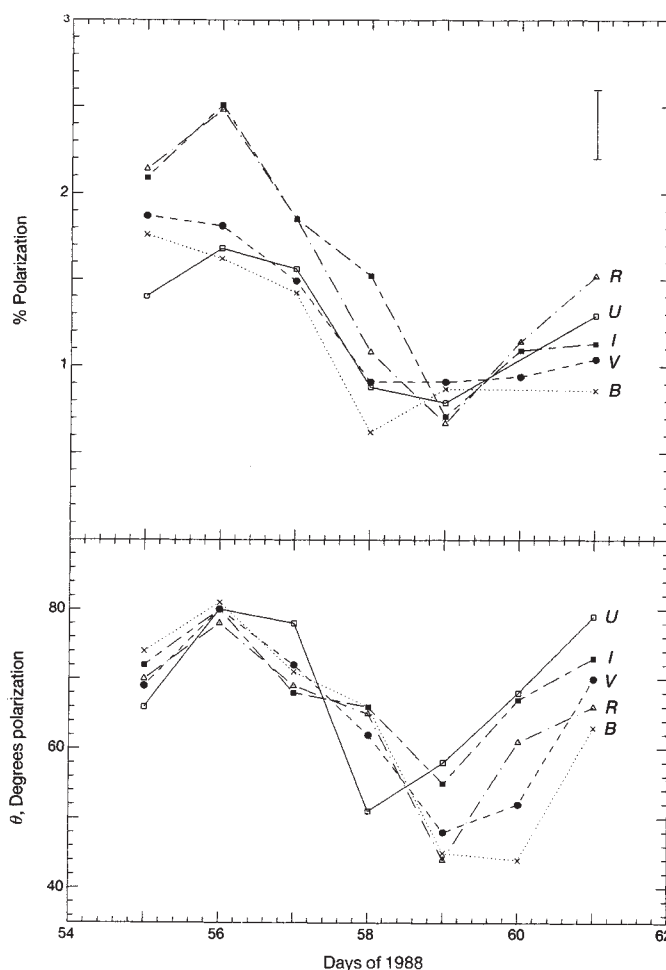


Fig. 2 Daily UBVR polarization measurements (%) and position angles, starting on 24 February.

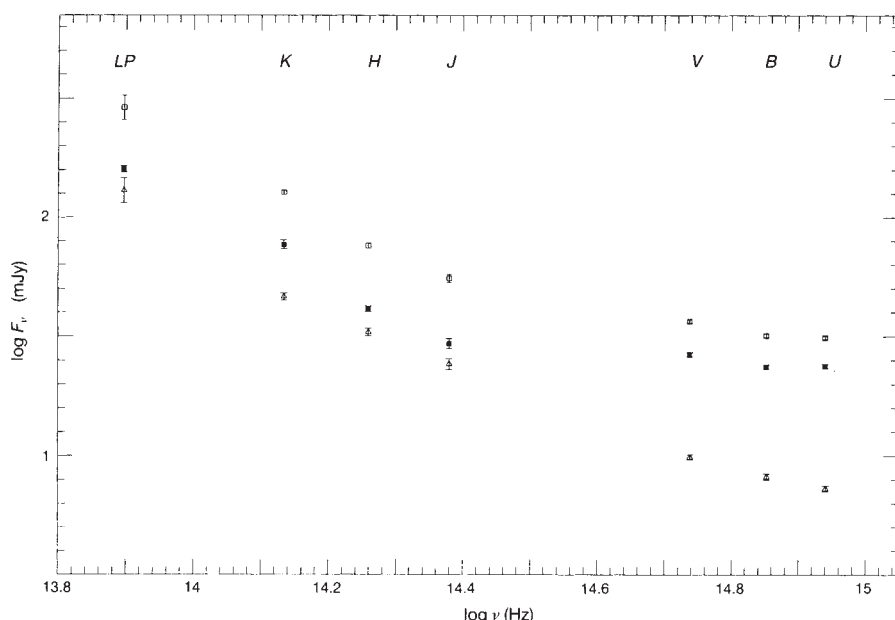


Fig. 3 Optical and infrared spectra for the quiescent emission (July 1987) (filled squares), for 10 March 1988 (open squares) and for the resulting flare spectrum (10 March 1988 minus July 1987) (open triangles). Frequency ν in Hz, F_ν in mJy.

data from June 1987 (assumed to represent the quiescent emission) and the difference between the two data sets, corresponding to the spectrum of the flare. In contrast to the quiescent spectrum and, to a lesser extent, the 10 March spectrum, the U - K energy distribution of the flare can be reasonably well represented by a single power law of slope $\alpha = 1.02$ ($F_\nu \propto \nu^{-\alpha}$).

Spectral evolution around the flare on 9 March can be seen from the infrared and optical slopes given in Table 1. We note that the spectral evolution of the U through K waveband during the period of maximum flare development of 8-13 March indicates a flattening of the infrared spectrum as the infrared flux reaches its maximum followed by a steepening as the flux decreases, and furthermore, the optical spectral index shows a rapid steepening from 10 March onward. The flattening expected as the flux rose to a maximum was not detected as a result of a lack of optical observations on 8 and 9 March.

In deriving emission parameters it is important to separate the effects of the flare from underlying 'quiescent' components. We follow Robson *et al.*³ in identifying a steady component in the near-infrared domain (1-3.8 μm), the origin of which is still unclear and which seems not to have fallen below a certain level over the past few years. An analysis of the long-term infrared variations is in preparation. This steady component is subtracted from the observed fluxes to obtain the flare emission. It is not clear whether at longer wavelengths the flare component can be separated from the quiescent emission in an additive manner, because the origin of both is thought to be synchrotron emission (see below) and therefore the same electrons can be partaking in both the quiescent and the flare emission. In addition, millimetre-to-infrared observations in 1986 showed that this component decreased by a large factor below its quiescent level³.

The spectral behaviour observed around 9 March is precisely what one expects for synchrotron cooling close to the cut-off frequency. The cooling-time measurements (about two days, Fig. 1b) therefore provide a direct measurement of the magnetic field. We can calculate this magnetic field using the relations $E/P = 7.6 \times 10^8 / (\gamma B^2)$ seconds and $\gamma^{1/2} = 2.3 \times 10^7 \nu/B$, where E is the energy of the electrons, γ their relativistic Lorentz factor, P the synchrotron power emitted, ν the characteristic frequency of the emitted radiation (in Hz) and B the magnetic field (in G). For $\nu = 2.5 \times 10^{14}$ Hz this gives $B = 0.7$ G and $\gamma = 9.1 \times 10^3$. We also note that the synchrotron cooling time is proportional to $\nu^{-1/2}$ and so a cooling time of one day at 1 μm leads to an expected cooling time of ~ 30 days at 1 mm.

The high polarization observed in 3C273 during late February and varying on a daily timescale (Fig. 2) is similar to the behaviour of blazars¹⁰. The polarization increases from U to I , which we interpret as the increasing dominance of the synchrotron component towards the infrared. We note, however, that the 10 March spectrum (Fig. 3) shows hardening (flattening) to higher frequencies, which is opposite to what is seen in blazars¹¹.

The brightness temperature of the flare on 9 March is 6.4×10^6 K at J for a size of one light day. Synchrotron self-Compton theory indicates that the self-Compton X-ray flux from a region with a brightness temperature of $\sim 10^7$ K will be small¹², which indicates, in turn, that the observed infrared (and optical) data does not imply relativistic beaming. However, data obtained on 23 February shows a steep infrared (1-3.8- μm) flare spectrum of slope 1.31. The 20- μm flux indicates a brightness temperature of $10^{13}/R^2$ where R is the source size in light days. Brightness temperatures exceeding 10^{12} K are excluded by synchrotron self-Compton theory for unbeamed sources¹², so beaming is required if the source size is less than three light days. Unfortunately, the sampling of the light curve at this epoch is insufficient to estimate such a small source size. But should the event of 10 March have had a spectrum similar up to 10 μm to that of 23 February, then beaming is definitely required, as this latter event was clearly more powerful and from a source with a size of the order of one light day. Since neither the wavelength of the maximum flare emission nor the flux at this maximum could be measured, we cannot make a quantitative estimate of the emitted X-ray flux.

It is interesting to note that the spectrum of the last optical maximum (5 April) is much harder ($\alpha = -0.02$, flare component) than that of 9-10 March. It might therefore be expected that this latter flare was not as important in the infrared. Fig. 1b tends to confirm this, indicating either an unusual electron distribution or a non-synchrotron origin.

We conclude that flaring events are much faster and more structured than previously accepted, even for sources as luminous as 3C273. The wealth of structure seen in these optical and infrared flares casts some doubt on previous descriptions of flares, which were based on sparsely sampled data. The fast flares observed in 3C273 require luminosity variations of 10^7 solar luminosities per second (without beaming), implying an unusual and powerful switch by any standards.

We thank many observers who have taken data on our behalf

and generously given their observing time for this programme, and ESO and the JAC for their rapid response and efficient execution of requests for urgent observations. Part of the infrared data were obtained as UKIRT Service observations. T.J.-L.C. acknowledges an SERC visiting fellowship to Lancashire Polytechnic where part of this work was done. T.J.-L.C. is affiliated to the Astrophysics Division, Space Science Department of the European Space Agency.

Received 1 July; accepted 10 August 1988.

1. Courvoisier, T. J.-L. *et al. Astr. Astrophys.* **176**, 197–209 (1987).
2. Ulrich, M.-H., Courvoisier, T. J.-L. & Wamsteker, W. *Astr. Astrophys.* (in the press).
3. Robson, E. I. *et al. Nature* **323**, 134–136 (1986).
4. Robson, E. I. *et al. Nature* **305**, 194–196 (1983).
5. Marscher, A. P. & Gear, W. K. *Astrophys. J.* **298**, 114–125 (1985).
6. Marshall, N., Warwick, R. S. & Pounds, K. A. *Mon. Not. R. astr. Soc.* **194**, 897–1002 (1981).
7. Sherwood, W. A., Kreysa, E., Gemuend, H.-P. & Biermann, P. *Astr. Astrophys.* **175**, L5–L6 (1983).
8. Rufener, F. & Nicolet, B. *Astr. Astrophys.* (in the press).
9. Tokunaga, A. *The NASA Infrared Telescope Facility Photometry Manual* (1986).
10. Brindle, C., Hough, J. H., Bailey, J. A., Axon, D. J. & Hyland, A. R. *Mon. Not. R. astr. Soc.* **221**, 739–768 (1986).
11. Brown, L. M. J. *et al. Astrophys. J.* (in the press).
12. Burbidge, G. R., Jones, T. W. & O'Dell, S. L. *Astrophys. J.* **193**, 43–54 (1974).

Future CFC concentrations under the Montreal Protocol and their greenhouse-effect implications

T. M. L. Wigley

Climatic Research Unit, University of East Anglia,
Norwich NR4 7TJ, UK

The Montreal Protocol on Substances that Deplete the Ozone Layer¹ is an agreement to reduce production and emissions of chlorofluorocarbons (CFCs) motivated primarily by the need to protect the Earth's ozone layer. The possible effects of CFCs on global climate were noted as an additional cause for concern. Here, the effect of the protocol on future concentrations of CFC-11 (CFCl_3) and CFC-12 (CF_2Cl_2) is examined, and the greenhouse-effect implications are evaluated. The protocol reduces the 1986–2030 warming commitment attributable to CFCs by a factor of three to seven, but high CFC concentrations (up to four times present levels for CFC-12) would still occur eventually unless further restrictions are imposed.

Global-mean tropospheric concentrations of CFC-11 and CFC-12 (respectively ~ 230 parts per 10^{12} by volume (p.p.t.v.) and ~ 395 p.p.t.v. in 1986^{2,3}) are increasing rapidly (by $\sim 6\%$ per year over 1977–86). There are a number of approaches to the problem of modelling these concentration changes. The more fundamental approach, which is necessary if information about the chemical consequences of CFC increases is required, is to use an appropriate chemical/transport model⁴. Such models, however, require as inputs emission data that are not directly available and which must be therefore estimated using production figures. A second approach is to use a much simpler representation of the chemistry, as part of a model which relies on production data as its basic input. This is the approach adopted here.

I have used a simple two-box model comprising an atmosphere box and a delayed-release box, the latter accounting for the fraction of production that does not immediately escape into the atmosphere. The model equations are

$$\frac{dM}{dt} = P_a + \alpha P_{na} + \beta D - M/\tau \quad (1)$$

$$\frac{dD}{dt} = (1 - \alpha)P_{na} - \beta D \quad (2)$$

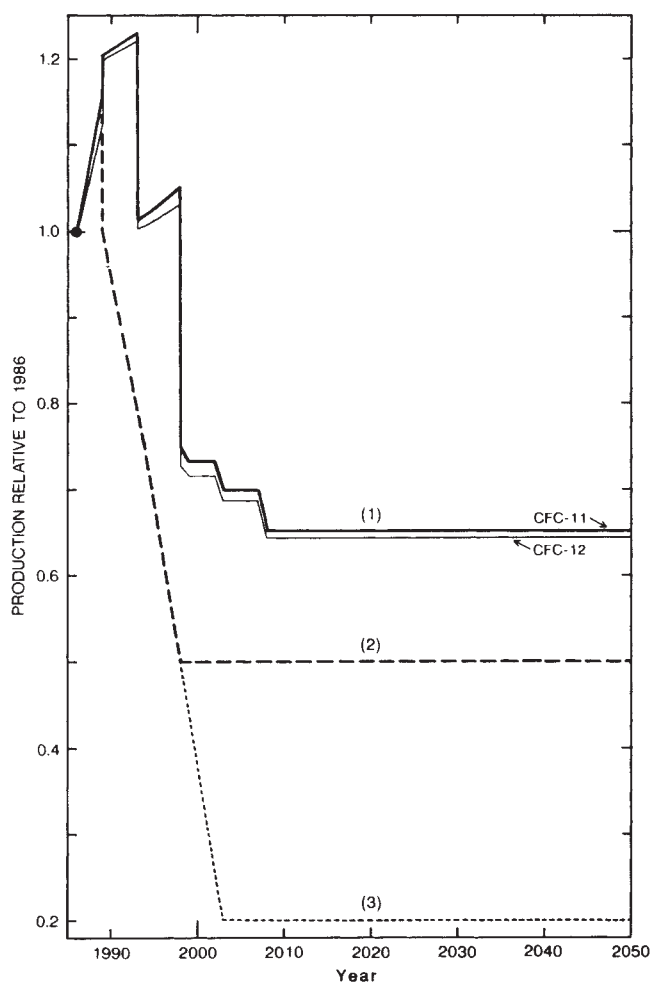


Fig. 1 Global CFC-production scenarios compatible with the Montreal Protocol. Production values are given relative to the 1986 values of $384.7 \text{ ktonne yr}^{-1}$ for CFC-11 and $492.5 \text{ ktonne yr}^{-1}$ for CFC-12. In all scenarios, production is assumed to grow over 1986–89 at 1981–86 growth rates, namely 5% per year for CFC-11 and 4% per year for CFC-12. Scenario (1) assumes that countries accounting for 90% of global production sign the protocol, that they use $1.1P_0$ (or $1.15P_0$ after 1998) as the base level, further inflated by 10% to account for facilities currently under construction. The remaining 10% of producers are developing countries, for which I assume a 10-yr delay before complying, with recent production growth rates assumed to hold until full compliance in 1999. Scenario (2) assumes that 100% of producers adopt the protocol, all using a base level of P_0 , and reducing production linearly from P_0 in mid-1989 to $0.8P_0$ in mid-1993 and then to $0.5P_0$ in mid-1998. Scenario (3) follows (2) to 1998 and then reduces production to $0.2P_0$ by 2003.

where P_a/P_{na} are the aerosol/non-aerosol production rates, M and D are masses in the atmospheric and delay boxes, α is the fraction of non-aerosols released directly to the atmosphere (the rest goes into the delay box), β is the annual fraction of the delay box that leaks into the atmosphere and τ is the lifetime.

Given production scenarios implied by the Montreal Protocol, equations (1) and (2) can be solved for future atmospheric masses and, hence, concentrations (C). This requires as input values for α , β and τ , the fraction of total future production (P) which is in the aerosol category, and initial (1986) values for M and D . As the protocol defines all future production in terms of the global value for 1986, this must be known also.

Suitable values for α and β are $\alpha = 0.25$, $\beta = 0.15$ for CFC-11 and $\alpha = 0.15$, $\beta = 0.25$ for CFC-12 (obtained by fitting the model to 1976–86 data). Estimates for lifetimes are given in ref. 5 as follows: CFC-11, $\tau = 75 \text{ yr}$ (range of uncertainty 58–107 yr);

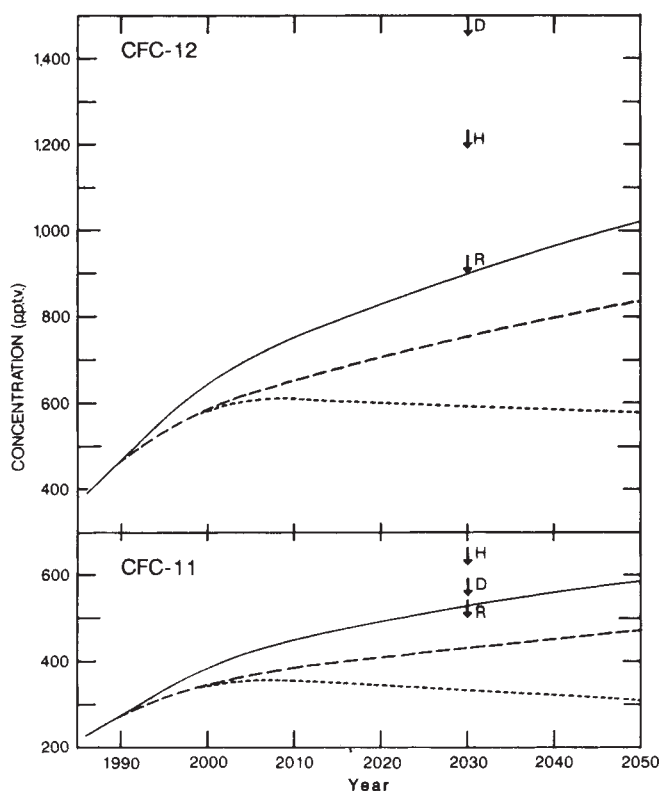


Fig. 2 Future concentrations of CFC-11 and CFC-12 (p.p.t.v.) for the production scenarios given in Fig. 1. The lowest curve corresponds to twenty-first century production levels well below those required by the protocol. The arrows marked H, D and R indicate the lower-limit pre-protocol estimates of Hammitt *et al.*⁷, Dickinson and Cicerone⁶ and Ramanathan *et al.*¹². For CFC-11 the concentration range for 2030 is (scenario (3)/(2)/(1)) 332/432/526 p.p.t.v., compared with earlier best estimates/ranges of 1,100/500–2,000 p.p.t.v. (ref. 12), 1,400/550–1,600 p.p.t.v. (ref. 6) and 890/620–1,330 p.p.t.v. (ref. 7). For CFC-12, the present estimated range of values for 2030 is 592/753/900 p.p.t.v., compared with 1,800/900–3,500 p.p.t.v. (ref. 12), 2,050/1,450–2,500 p.p.t.v. (ref. 6) and 1,690/1,190–2,510 p.p.t.v. (ref. 7).

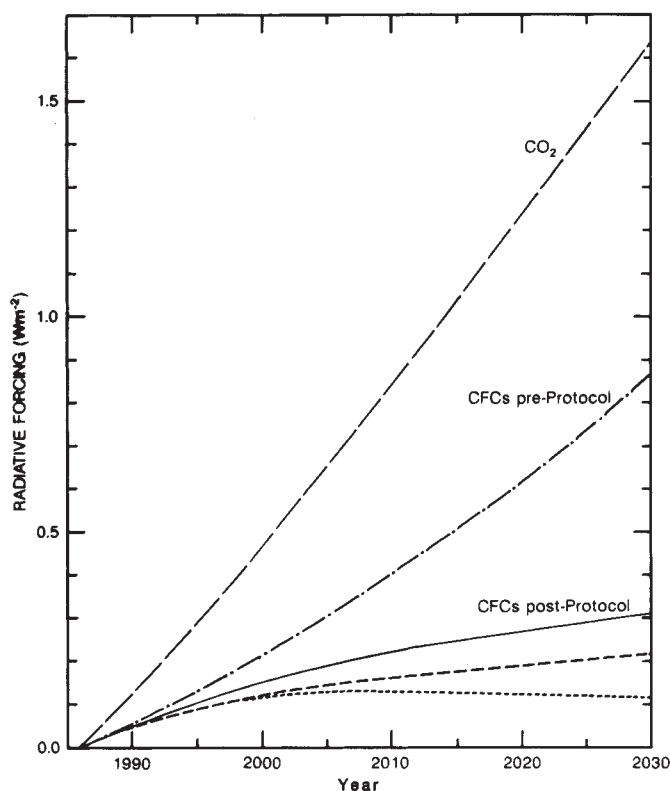


Fig. 3 Radiative forcing changes after 1986 for expected CO₂ increases, and for both the CFC increases expected before the Montreal Protocol and for the CFC increases likely to occur in response to the protocol. The lowest curve corresponds to twenty-first century production levels well below those required by the protocol. CO₂ projections correspond to a concentration of 450 parts per million by volume in 2030, the best estimate of Ramanathan *et al.*¹² and lying in the low end of the range given by Trabalka *et al.*¹³. Pre-protocol CFC projections correspond to concentrations of 1,050 p.p.t.v. for CFC-11 and 1,790 p.p.t.v. for CFC-12 in 2030, which are close to the means of the best estimates in refs 6, 7 and 12. Refs 3 and 14 give further details.

CFC-12, $\tau = 111$ yr (65–400 yr). Dickinson and Cicerone⁶ give values of 60 yr (50–80 yr) for CFC-11 and 100 yr (90–130 yr) for CFC-12, and Hammitt *et al.*⁷ give 76.5 yr for CFC-11 and 138.8 yr for CFC-12.

Initial values of M can be obtained from the tropospheric concentrations ($C_{z=0}$), provided the vertical (z) concentration profile is known. Based on Hough *et al.*⁸, the column-mean concentration ($\bar{C}$) for both CFC-11 and CFC-12 is given by $\bar{C} \approx 0.97 C_{z=0}$. Hence $M = a C_{z=0}$ where $a = 24.25$ for CFC-11 and 21.32 for CFC-12 (M in ktonne, C in p.p.t.v.). Initial values of D can be estimated using figures from the Fluorocarbon Program Panel of the Chemical Manufacturers Association (FPPCMA)⁹, scaled to account for non-reporting countries by assuming that the ratio of D to accumulated total production is constant. This gives global values for 1986 of 947.5 ktonne for CFC-11 and 862.6 ktonne for CFC-12. Future values of the aerosol fraction of production (P_a/P) can be estimated from FPPCMA figures for 1976–86⁹. These show a steady decline in P_a/P for CFC-11 and CFC-12 from almost 60% in 1976 to around 30% in 1986. Based on this downward trend, I have assumed

$$P_a = P[0.30 \exp(-0.05t)] \quad (3)$$

where t is in years after 1986.

1986 production values can only be estimated because not all producers supply data to the FPPCMA. The main data gap is the USSR, for which information is available only for 1968–1975¹⁰. Over this period USSR production increased by 27% per yr for CFC-11 and 18% per yr for CFC-12, but it is unlikely that these rates have been maintained. To estimate 1986 USSR levels, I assume post-1975 growth rates that are constrained to give a 1982 production figure for CFC-11 plus CFC-12 consistent with ref. 11 (76 ktonne per yr), with rates in the ratio 27:18. To obtain global production figures I have increased the USSR figures by 1.15, in accord with the estimate for eastern European production¹¹. This gives 1986 global production of 384.7 ktonne yr⁻¹ for CFC-11 and 492.5 ktonne yr⁻¹ for CFC-12, as compared with production for reporting countries of 350.1 ktonne yr⁻¹ for CFC-11 and 398.4 ktonne yr⁻¹ for CFC-12 (ref. 9). Thus, I estimate that reporting countries contributed 85% of the global production in 1986, in excellent agreement with the estimate of 86% by Hammitt *et al.*⁷.

The next step is to determine a range of future production scenarios consistent with the protocol. This requires a cut-back to 1986 production levels (P_0) in mid-1989, to 0.8 P_0 in mid-1993, and to 0.5 P_0 in mid-1998. Flexibility in these figures arises in the following ways: a base level of 1.1 P_0 (or 1.15 P_0 after 1998) may be used to allow for exports to low-level producers; this

Table 1 Model sensitivity to α , β and τ

α	β	τ	1990	2000	2010	2020	2030	2040	2050
0.15	0.25	111	464	587	651	704	753	798	839
0.15*	0.25*	111*	465	592	657	712	761	806	847
0.075	0.25	111	460	585	649	702	751	796	837
0.30	0.25	111	470	593	656	709	758	802	843
0.15	0.125	111	445	552	620	677	727	774	817
0.15	0.50	111	484	606	665	717	765	810	850
0.15	0.25	150	468	603	680	747	809	868	924
0.15	0.25	80	458	565	611	646	678	706	731

Values shown are CFC-12 concentration in p.p.t.v., for production scenario (2). Results in row 2(*) are for the same parameters as row 1 but with the aerosol fraction held constant at 0.3 rather than decaying according to equation (3).

base level may be further increased to incorporate production facilities presently under construction; developing countries may delay compliance with the protocol by up to 10 years; and the protocol may come into force even if countries accounting for only two-thirds of the estimated 1986 global consumption sign the agreement. Based on these details, three scenarios were constructed (see Fig. 1).

Future concentrations under the three models are shown in Fig. 2. These are subject to a considerable range of uncertainty, but this range is almost an order of magnitude less than it was before the Montreal Protocol. More importantly, the protocol implies much smaller concentration increases than earlier estimates. Even scenario (1) has values less than most previous lower-limit estimates.

In scenarios (1) and (2), future concentrations continue to rise beyond 2050 to asymptotic concentration limits of 770 p.p.t.v. and 590 p.p.t.v., respectively, for CFC-11 and 1,650 p.p.t.v. and 1,280 p.p.t.v. for CFC-12. For scenario (3), which goes beyond the present protocol in assuming an eventual cut-back in production to 20% of the 1986 level, concentrations reach their maxima early in the twenty-first century and then slowly decline to asymptotic values of 240 p.p.t.v. (CFC-11) and 510 p.p.t.v. (CFC-12). Production limits of ~19% (CFC-11) and 15% (CFC-12) of 1986 values would be required to stabilize atmospheric concentrations at levels near those prevailing today.

To assess the effect of uncertainties in the various model parameters on these results, sensitivity studies were carried out using scenario (2) and varying α , β , τ and the ratio P_a/P . Details for CFC-12 are shown in Table 1. Only the lifetime uncertainty affects the results by more than a few per cent. Changing τ by $\pm 30\%$ alters the projected concentrations in 2030 by around $\pm 10\%$. (Because the asymptotic concentration for constant production rate is proportional to τ , the sensitivity to $\pm 30\%$ variations in τ would eventually be $\pm 30\%$.)

The greenhouse-effect implications of the above three scenarios determined using radiative transfer results from Ramanathan *et al.*¹² For CFC-11 and CFC-12, the direct radiative forcing perturbations (ΔQ) depend linearly on

concentration:

$$\text{CFC-11: } \Delta Q = 0.00027 C \text{ (CFC-11)}$$

$$\text{CFC-12: } \Delta Q = 0.00031 C \text{ (CFC-12)}$$

(ΔQ in W m^{-2} , C in p.p.t.v.). To account for CFCs other than CFC-11 and CFC-12, the (CFC-11 + CFC-12) forcing has been increased by 32%. This accords with changes up to 1985³ and assumes future changes in the production of all CFCs that parallel those for CFC-11 and CFC-12.

The greenhouse implications are illustrated in Fig. 3. The expected forcing contribution from CO_2 increases and an intermediate CFC-forcing projection estimated before the Montreal Protocol are shown for comparison, both from ref. 3. Before the protocol, the expected future forcing contribution from CFC increases was ~50% of that arising from CO_2 increases. Now, future CFC-forcing between 1986 and 2030 amounts to only between 7% and 19% of the CO_2 -forcing or to between 14% and 36% of the pre-protocol CFC-forcing. As the global-mean warming commitment is directly proportional to the forcing, this means that the protocol will reduce the CFC contribution to this commitment by a factor of three to seven. Another way to view the impact of the protocol is in terms of the delaying effect it will have on future greenhouse-gas-forcing changes. In ref. 14 I have shown that, using best possible (pre-protocol) estimates of future greenhouse-gas concentration changes, an equivalent doubling of the pre-industrial CO_2 concentration is reached in 2019. If the results for scenario (2) are used, the doubling date is delayed until 2027.

It is possible that the results presented above may be underestimated if, for example, alternative CFCs (such as HCFC-22) were used to replace CFC-11 and CFC-12. As an extreme example, I have supposed that the pre- and post-protocol differences in (CFC-11 + CFC-12) production result entirely from increases in HCFC-22 (CHClF_2) production. Using $\tau = 20$ yr (ref. 12) and $\alpha = 0.2$, $\beta = 0.2$, the model gives an increase in HCFC-22 concentration from 77 p.p.t.v. in 1986 to 1,670 p.p.t.v. in 2030. The implied radiative forcing change, using $\Delta Q = 0.0010 C$ (HCFC-22) from ref. 12, is 0.16 W m^{-2} . This should be compared with the differences between the pre-protocol change of 0.86 W m^{-2} and the post-protocol changes of 0.12, 0.22 and 0.31 W m^{-2} for models (1), (2) and (3) respectively. Substitution by HCFC-22 would lead to a substantial reduction in future forcing, even in the most unlikely case of full substitution. Other substitutes may, however, be less benign.

The emphasis in this paper has been on future CFC-related greenhouse effects. In this regard, the Montreal Protocol represents a significant step forward. But one should not lose sight of the fact that climate is a secondary consideration in limiting CFC production. Under the present protocol terms, large increases in CFC concentrations can still occur and these increases can only serve to further exacerbate the ozone depletion problem.

This work was supported by the US Department of Energy, Carbon Dioxide Research Division, and the Commission of the European Communities Climatology programme.

Received 9 May; accepted 23 August 1988.

1. United Nations Environment Programme, *Montreal Protocol on Substances that Deplete the Ozone Layer* (United Nations Environment Programme, Nairobi, 1987).
2. Rasmussen, R. A. & Khalil, M. A. K. *Science* **232**, 1623–1624 (1986).
3. Wigley, T. M. L. *Clim. Monit.* **16**, 14–28 (1987).
4. Brühl, C. & Crutzen, P. J. *Climate Dyn.* **2**, 173–203 (1988).
5. Gammon, R. H. & Wofsy, S. C. in *Atmospheric Ozone 1985: Assessment of Our Understanding of the Processes Controlling its Present Distribution and Change*, Global Ozone Research and Monitoring Project, Report No. 16 (World Meteorological Organisation, Geneva, 1985).
6. Dickinson, R. E. & Cicerone, R. J. *Nature* **319**, 109–115 (1986).
7. Hammit, J. K. *et al. Nature* **330**, 711–716 (1987).

8. Hough, A. M., Penkett, S. A., Eggleton, A. E. J., Schmidt, U. & Knapska, D. A *Single Dimension Modelling Study of the Vertical Profiles of Halocarbons and Hydrocarbons at 44° N*, AERE R 12588, (UK Atomic Energy Authority, Harwell, 1987).
9. Chemical Manufacturers Association *Production, Sales, and Calculated Release of CFC-11 and CFC-12 through 1986* (Chemical Manufacturers Association, Washington, DC, 1987).
10. Borisenkov, E. P. & Kazakov, Y. E. *Trudy Gl. Geofiz. Obs.* **438**, 62–74 (1980).
11. Gamlen, P. H., Lane, B. C. & Midgley, P. M. *Atmos. Envir.* **20**, 1077–1085 (1986).
12. Ramanathan, V., Cicerone, R. J., Singh, H. B. & Kiehl, J. T. *J. geophys. Res.* **90**, 5547–5566 (1985).
13. Trabalka, J. R., Edmonds, J. A., Reilly, J. M., Gardner, R. H. & Voorhes, L. D. in *Atmospheric Carbon Dioxide and the Global Carbon Cycle* (ed. Trabalka, J. R.) 247–287 (US Dept of Energy, Washington, DC, 1985).
14. Wigley, T. M. L. *Geophys. Res. Lett.* **14**, 1135–1138 (1987).

Simultaneous observations of ammonia in the atmosphere and ocean

Patricia K. Quinn*, Robert J. Charlson†
& Timothy S. Bates‡

* Department of Chemistry, BG-10, University of Washington, Seattle, Washington 98195, USA

† Department of Atmospheric Sciences, AK-40, University of Washington, Seattle, Washington 98195, USA

‡ National Oceanic and Atmospheric Administration/Pacific Marine Environmental Laboratory, 7600 Sand Point Way NE, Seattle, Washington 98115, USA

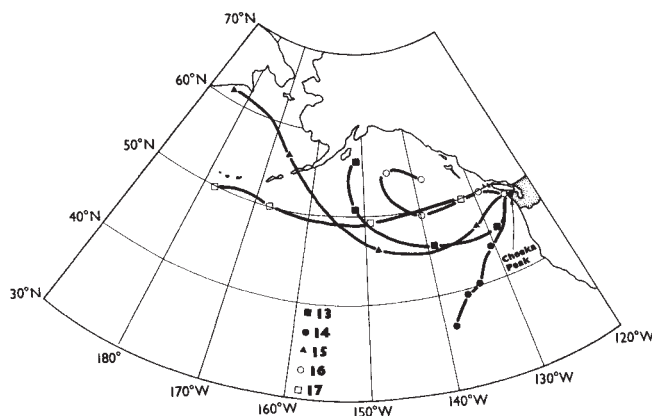


Fig. 1 Study area along the coast of Washington State. Samples were collected from the NOAA ship *McArthur* (5–200 km from shore) and on top of Cheeka Peak at the northernmost tip of Washington State. Also shown are the calculated four-day isentropic air mass trajectories* for 13–17 May 1987.

Ammonia, being the dominant base other than sea salt in the remote marine troposphere, plays an important role in the acid-base chemistry of the atmosphere^{1–3}. An understanding of the cycling of ammonia through the marine environment has, until now, been hampered by the lack of concurrent measurements of key acidic and basic atmospheric and seawater species. Reported here are the results of simultaneous concentration measurements of these species during May 1987 in the coastal north-east Pacific Ocean environment. Gas (g) and particulate (p) phase data suggest that low concentrations of NH_3 (g) in the marine atmosphere lead to partly neutralized H_2SO_4 in aerosol particles and in rainwater. Gas-phase concentrations combined with seawater concentration measurements indicate that for the region and time period studied the ocean was a local source of atmospheric ammonia. These data, combined in a simple box model, suggest a marine boundary layer residence time of 3.6 h and 22 h for NH_3 (g) and NH_4^+ (p), respectively.

Samples were collected from 8 to 22 May 1987 at the coastal mountaintop site on Cheeka Peak (48°18' N; 124°37' W; altitude 480 m) on the northernmost tip of Washington state and aboard the NOAA ship, *McArthur*, in the coastal waters of the north-east Pacific Ocean (Fig. 1). During this period the biological productivity in the surface ocean waters was seasonably high with primary productivity rates of 12 to 450 $\text{mg C m}^{-3} \text{ d}^{-1}$. The phytoplankton standing stock, as measured by chlorophyll 'a', ranged from 1 to 15 mg m^{-3} . The species sampled were atmospheric gas phase ammonia, NH_3 (g), atmospheric particulate phase ammonium and non-seasalt sulphate, NH_4^+ (p) and n.s.s. SO_4^{2-} (p) respectively, dissolved unionized and ionized ammonia in sea water, NH_3 (s) and NH_4^+ (s), and NH_3 in equilibrium with sea water.

Atmospheric samples were collected 7 m above the land surface at Cheeka Peak. The shipboard samples were collected from a boom extending 6 m upwind of the ship and 10 m above the sea surface. Samples were taken only during periods of SW to NW on-shore winds and when the total particle (CN) count was less than 1,000 cm^{-3} . This procedure coupled with the use of calculated air mass trajectories⁴ (Fig. 1), ensured the sampling of purely remote marine air.

A tandem sampling system was used which consisted of a cyclone separator (50 per cent collection efficiency at 0.9 μm for NaCl particles) followed by a 47-mm Millipore Teflon filter (1.0- μm pore size) for the collection of particulate species and four oxalic-acid-impregnated Whatman 41 filters for the collection of NH_3 (g). The cyclone fraction was discarded. The four Whatman filters were used in series to ensure a high collection efficiency for NH_3 (g). Laboratory tests have shown the NH_3 (g) collection efficiency of this system to be nearly 100% (P. K. Quinn and T. S. Bates, manuscript in preparation). A second stacked filter pack was mounted alongside each sample to serve as a sampling and analytical blank (sample flow rate was 52 standard l m^{-1} while blank flow rate was 0 standard l m^{-1}). The NH_4^+ (p) blank concentration was typically below the detection limit of 0.1 $\mu\text{mol l}^{-1}$, the mean n.s.s. SO_4^{2-} (p) blank concentra-

tion was 11% of the sample concentration and the mean NH_3 (g) blank concentration was 30% of the sample concentration.

Bulk seawater samples were taken from a depth of 5 m for the determination of the NH_4^+ (s) + NH_3 (s) concentration. An equilibrator, consisting of a closed volume of air in contact with a large volume of continually replenished sea water⁵, was used to measure the saturation concentration of ammonia, $(\text{NH}_3)_{\text{sg}}$, relative to its concentration in ocean surface waters. Measurements of $[\text{NH}_4^+] + [\text{NH}_3]_{\text{s}}$ and $(\text{NH}_3)_{\text{sg}}$ showed that the equilibrator functioned within a factor of three of Henry's Law. (Round brackets indicate concentrations in the atmosphere, that is, in the gas and particle phases. Square brackets indicate concentrations in the liquid phase, that is, in sea water and rainwater.) But this calculation was based on a measured Henry's Law constant⁶, K_{HN} , for a low ionic strength solution measured at 25 °C and corrected to the equilibrator sea water temperature of 11 °C. A measured K_{HN} for NH_3 in sea water at the appropriate temperature would improve these calculations. The NH_3 blank concentration for this system (equilibrator tubing, pump and filter pack, excluding the sea water) was typically 20% of the sample value. The bulk seawater samples were taken from the same line that served as the sea water supply for the equilibrator.

The phenylhypochlorite colorimetric technique⁷ was used for all ammonia analyses. Ion chromatography was used for the SO_4^{2-} and Na^+ determinations. All ammonia samples were analysed within 6 h of collection while the SO_4^{2-} and Na^+ filter samples were extracted with methanol and water and frozen for later analysis. All sample handling was carried out in an NH_3 and SO_2 free glove box, both in the laboratory and field.

The mean and sample standard deviation of the measured shipboard and Cheeka Peak $(\text{NH}_3)_{\text{g}}$ were $1.3 \pm 0.7 \text{ nmol m}^{-3}$ (number of samples $n = 14$) and $0.76 \pm 0.35 \text{ nmol m}^{-3}$ ($n = 8$), respectively, with an average difference between duplicate samples of 41%. The range of concentrations measured here encompasses the mean value of 3.5 nmol m^{-3} measured in maritime air at Cape Grim (40°41' S; 144°41' E) (ref. 8). The shipboard and Cheeka Peak concentrations of NH_4^+ (p) were $5.9 \pm 2.4 \text{ nmol m}^{-3}$ ($n = 13$) and $6.1 \pm 3.2 \text{ nmol m}^{-3}$ ($n = 8$), respectively, with an average difference between duplicate samples of 33%. These values fall within the range of $(\text{NH}_4^+)_{\text{p}}$, 0.5–15 nmol m^{-3} , found over the Pacific Ocean by Parungo *et al.*⁹. Non-seasalt $(\text{SO}_4^{2-})_{\text{p}}$ were calculated from measured $(\text{Na}^+)_{\text{p}}$. The shipboard and Cheeka Peak n.s.s. $(\text{SO}_4^{2-})_{\text{p}}$ were $5.1 \pm 2.6 \text{ nmol m}^{-3}$ ($n = 7$) and $4.6 \pm 2.0 \text{ nmol m}^{-3}$ ($n = 4$), respectively. The seasalt $(\text{SO}_4^{2-})_{\text{p}}$ was always less than 8% of the total $(\text{SO}_4^{2-})_{\text{p}}$.

The NH_4^+ (p) to n.s.s. SO_4^{2-} (p) molar ratio measured on the ship was 1.3 ± 0.7 while that at Cheeka Peak was 1.3 ± 0.3 (Fig. 2).

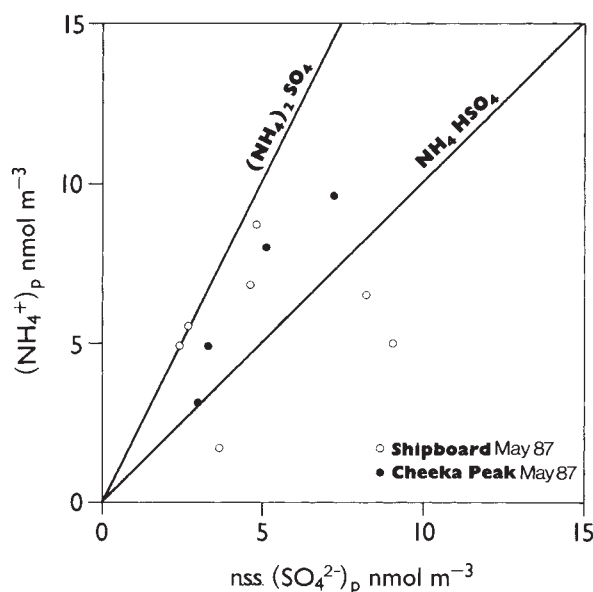


Fig. 2 Particulate n.s.s. $(\text{SO}_4^{2-})_p$ versus $(\text{NH}_4^+)_p$ for air masses over the NE Pacific Ocean during May 1987. Shipboard and Cheeka Peak samples have a correlation coefficient of 0.83 and 0.89, respectively. Samples falling on the line marked $(\text{NH}_4)_2\text{SO}_4$ have an NH_4^+ (p) to n.s.s. SO_4^{2-} (p) molar ratio of 2 while those on the line marked NH_4HSO_4 have a ratio of 1.

Fully neutralized H_2SO_4 aerosol particles would have an NH_4^+ (p) to n.s.s. SO_4^{2-} (p) molar ratio of 2. The measured molar ratio suggests a mean particulate chemical composition between NH_4HSO_4 and $(\text{NH}_4)_2\text{SO}_4$ and is consistent with the NH_4^+ (r) to n.s.s. SO_4^{2-} (r) molar ratio of ~ 1 found in remote marine rain^{2,10}. On the basis of these data, it appears that the low availability of NH_3 (g) in remote marine air leads to partially neutralized H_2SO_4 particulates. Since some of the partly neutralized H_2SO_4 aerosol particles act as cloud condensation nuclei¹¹ and are the main source of n.s.s. SO_4^{2-} in remote marine rainwater, low NH_3 (g) concentrations yield rainwater that is only 30–60% neutralized¹².

The flux of a gas-phase rate-controlled gas into or out of the ocean surface can be described by

$$F = k_g(c_{sg} - c_g) \quad (1)$$

where F is the flux, k_g is the gas phase transport coefficient of the gas, c_{sg} is the saturation concentration of the gas relative to its concentration in the surface ocean waters, and c_g is the bulk atmospheric gas concentration¹³. For ammonia, c_g is $(\text{NH}_3)_g$ and c_{sg} is $(\text{NH}_3)_{sg}$ and is related to $[\text{NH}_3]_s$, the dissolved unionized form of the gas in the ocean, through Henry's Law. This relation is given by

$$K_{hN} = (\text{NH}_3)_{sg} / [\text{NH}_3]_s \quad (2)$$

To date, all seawater measurements of ammonia species have yielded only the sum of the NH_4^+ (s) and NH_3 (s) concentrations. Since NH_3 (s) is the active species in the air/sea exchange of ammonia, it is necessary to know its saturation concentration relative to its concentration in ocean surface waters to obtain an accurate estimate of the flux of ammonia into or out of the ocean.

Bulk seawater, equilibrator, and gas phase atmospheric samples were collected simultaneously for analysis of $[\text{NH}_4^+]_s + [\text{NH}_3]_s$, $(\text{NH}_3)_{sg}$ and $(\text{NH}_3)_g$, respectively. The bulk seawater $([\text{NH}_4^+]_s + [\text{NH}_3]_s)$ was $0.45 \pm 0.35 \mu\text{mol l}^{-1}$ ($n = 48$). $(\text{NH}_3)_{sg}$ ranged from 4.9 to 22 nmol m^{-3} with a mean of 16 ± 7.1 ($n = 6$) while $(\text{NH}_3)_g$ ranged from 0.81 to 1.4 nmol m^{-3} with a mean of 0.97 ± 0.27 ($n = 10$).

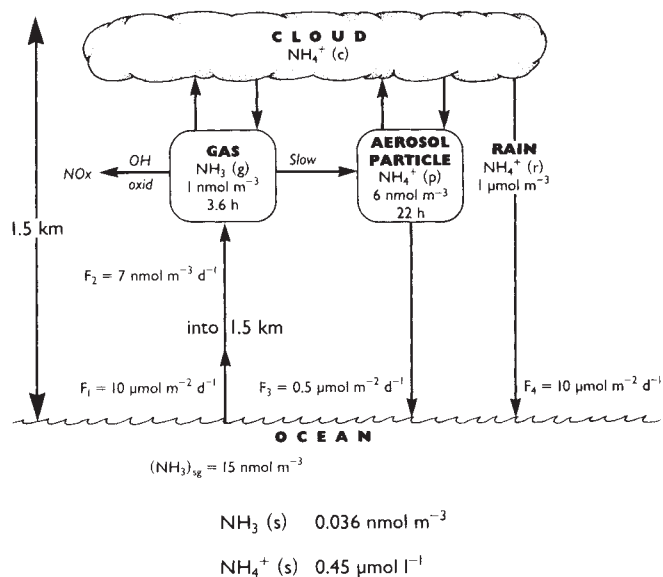


Fig. 3 Box model for ammonia species in the remote marine environment over the NE Pacific Ocean during May 1987. The average box height (1.5 km) was determined from radiosonde measurements. All concentrations were measured simultaneously. $F_1 = k_g(c_{sg} - c_g) = 10 \mu\text{mol m}^{-2} \text{d}^{-1}$. This flux of NH_3 (g) into a height of 1.5 km corresponds to an input rate F_2 of $7 \text{ nmol m}^{-3} \text{d}^{-1}$. F_3 is the dry deposition flux of NH_4^+ (p) using the given concentration of NH_4^+ (p) and a deposition velocity of 0.1 cm s^{-1} , which corresponds to a particle radius $< 0.5 \mu\text{m}$. F_4 is based on the given rainwater concentration, a total rainfall amount of 2.0 cm, and the time since the last rainfall, 2 days. The lifetimes of NH_3 (g) and NH_4^+ (p), t_1 and t_2 respectively, are calculated from the given concentrations and the value of F_2 .

$(\text{NH}_3)_{sg}$ were consistently an order of magnitude higher than $(\text{NH}_3)_g$ which, by use of equation (1), suggests that for the region studied the ocean was a net source of ammonia. The flux of ammonia out of the ocean calculated via equation (1) was found to range between 2.6 and $16 \mu\text{mol m}^{-2} \text{d}^{-1}$ with a mean of 10 ± 5 ($n = 6$). The transport coefficient used was that of H_2O , $3,000 \text{ cm h}^{-1}$, as both H_2O and NH_3 are gas-phase rate-controlled species and have similar molecular weights¹³. An empirically determined k_g for NH_3 would improve the accuracy of these calculations.

The flux derived from the measured $(\text{NH}_3)_{sg}$ and $(\text{NH}_3)_g$ values was compared to one calculated from bulk seawater measurements of $[\text{NH}_3]_s + [\text{NH}_4^+]_s$ as described by Ayers and Gras⁸. Based on the $([\text{NH}_3]_s + [\text{NH}_4^+]_s)$, F was calculated to be $3.1 \mu\text{mol m}^{-2} \text{d}^{-1}$. Although this flux is within the range of values calculated using the equilibrator values of $(\text{NH}_3)_{sg}$, it is a factor of three lower than the calculated mean flux of $10 \mu\text{mol m}^{-2} \text{d}^{-1}$. At least one difficulty in using $([\text{NH}_3]_s + [\text{NH}_4^+]_s)$ for the determination of $(\text{NH}_3)_{sg}$ is the lack of a measured K_{hN} for NH_3 in sea water. The equilibrator technique does not require the use of K_{hN} . Although it has not been used previously for NH_3 measurements it has been shown to work well with other gases (CO_2 , CH_4 , DMS) (ref. 5 and J. E. Johnson, unpublished data).

The NH_3 (g), NH_4^+ (p), NH_4^+ (s) and NH_3 (s) data are compared in a box model in Fig. 3. NH_3 is released from the ocean to the atmosphere with an average flux $F_1 = 10 \mu\text{mol m}^{-2} \text{d}^{-1}$. An average height of the well mixed marine boundary layer during this study, as determined from radiosonde measurements, was $\sim 1.5 \text{ km}$. The flux of $10 \mu\text{mol NH}_3 \text{ m}^{-2} \text{d}^{-1}$ corresponds to an input rate into this height, F_2 , of $7 \text{ nmol m}^{-3} \text{d}^{-1}$. The NH_3 (g) concentration measured in this study was $\sim 1 \text{ nmol m}^{-3}$. The majority of the NH_3 (g) reacts with H_2SO_4 aerosol particles or

is dissolved into cloud droplets. Since these two processes occur much faster than the reaction of NH_3 (g) with OH radical, only a small portion of the NH_3 is oxidized¹⁴⁻¹⁶. The measured average NH_4^+ (p) concentration in the remote marine troposphere was 6 nmol m^{-3} . There are no reported measurements to date of NH_4^+ in remote marine cloudwater.

Turnover times of $t_1 = 3.6 \text{ h}$ and $t_2 = 22 \text{ h}$ for NH_3 (g) and NH_4^+ (p), respectively, were calculated from average concentrations and F_2 . The short turnover time for NH_3 (g) reflects its rapid accretion onto the condensed phase during passage through a cloud and its subsequent inability to return to the gas phase after the cloud droplet evaporates. The time of 3.6 h is consistent with vertical mixing times of the marine planetary boundary layer¹⁴.

A dry deposition flux for NH_4^+ (p) (F_3) of $0.5 \mu\text{mol m}^{-2} \text{ d}^{-1}$ was calculated using the average $(\text{NH}_4^+)_p$ and a deposition velocity of 0.1 cm s^{-1} which corresponds to a particle radius of less than $0.5 \mu\text{m}$. The rainwater concentration of NH_4^+ was $1.0 \pm 0.1 \mu\text{mol l}^{-1}$ during one storm sampled aboard the ship *McArthur* on 13 May 1987. Based on a total rainfall amount of 2.0 cm and the time since the previous rainfall (two days), the wet depositional flux (F_4) was $10 \mu\text{mol m}^{-2} \text{ d}^{-1}$. This rainwater flux is consistent with the range of 2 to $14 \mu\text{mol m}^{-2} \text{ d}^{-1}$ measured over the Pacific Ocean² by Galloway. The agreement of the wet deposition flux of NH_4^+ , F_4 , with the measured flux out of the ocean, F_1 , ($10 \mu\text{mol m}^{-2} \text{ d}^{-1}$) indicates that NH_3 is conserved as it cycles through the marine environment.

We thank Dr J. Johnson, Dr D. Covert, Mr J. Zieman, Mr R.

Peterson and Mr E. Prestbo for sampling assistance, Dr C. Hahn for the air mass trajectories, the officers and crew of the NOAA ship *McArthur* for their assistance and cooperation, and Drs W. Asher, J. Galloway, R. Gammon, R. Harriss, J. Johnson and A. Langford for helpful comments. This work was supported by NSF and by NOAA's National Acid Precipitation Assessment Program (NOAA Pacific Marine Environmental Laboratory).

Received 7 April; accepted 25 July 1988.

1. Charlson, R. J. & Rodhe, H. *Nature* **295**, 683-685 (1982).
2. Galloway, J. N. *NATO: The Biogeochemical Cycling of Sulfur and Nitrogen in the Remote Atmosphere* 143-175 (Reidel, Boston, 1985).
3. Quinn, P. K., Charlson, R. J. & Zoller, W. H. *Tellus* **39B**, 413-425 (1987).
4. Haagensen, P. L. & Shapiro, M. A. *Isentropic Trajectories for Derivation of Objectively Analyzed Meteorological Parameters* NCAR Tech. Note 149TSTR (NCAR, Boulder, 1979).
5. Keeling, C. D., Rackstraw, N. W. & Waterman, L. S. *J. geophys. Res.* **70**, 6087-6102 (1965).
6. Stumm, W. & Morgan, J. J. *Aquatic Chemistry* 109 (Wiley, New York, 1981).
7. Solarzano, L. *Limnol. Oceanogr.* **14**, 799-801 (1969).
8. Ayers, G. P. & Gras, J. L. *Nature* **284**, 539-540 (1980).
9. Parungo, F. P., Nagamoto, C. T., Rosinski, J. & Haagensen, P. L. *J. atmos. Chem.* **4**, 199-226 (1986).
10. Vong, R. J., Hansson, H. C., Ross, H. B., Covert, D. S. & Charlson, R. J. *J. geophys. Res.* **93**, 1625-1637 (1988).
11. Pruppacher, H. R. & Klett, J. D. *Microphysics of Clouds and Precipitation* 228-229 (Reidel, Boston, 1978).
12. Vong, R. J., Hansson, H. C., Covert, D. S. & Charlson, R. J. *Geophys. Res. Lett.* **15**, 338-341 (1988).
13. Liss, P. S. & Slater, P. G. *Nature* **247**, 181-184 (1974).
14. Junge, C. *The Modification of Aerosol Size Distribution in the Atmosphere* US Department of Commerce/NBS/Institute for Applied Technology, AD 445 873 (1964).
15. Crutzen, P. J. *The Major Biogeochemical Cycles and Their Interactions* 84 (Wiley, New York, 1983).
16. Huntzicker, J. J., Cary, R. A. & Ling, C. *Envir. Sci. Technol.* **14**, 819-824 (1980).

Origin of technetium-99 and its use as a marine tracer

A. Aarkrog*, L. Carlsson†, Q. J. Chen*,
H. Dahlgaard*, E. Holm‡, L. Huynh-Ngoc‡,
L. H. Jensen*, S. P. Nielsen* & H. Nies§

* Risø National Laboratory, DK-4000 Roskilde, Denmark

† University of Lund, Radiation Physics Department,
S-22185 Lund, Sweden

‡ International Laboratory of Marine Radioactivity,
MC-98000, Monaco

§ German Hydrographic Institute, D-2000 Hamburg 4, FRG

The Chernobyl accident caused a variety of radionuclides to be released. The most abundant of the longer-lived nuclides were the two caesium isotopes ^{137}Cs and ^{134}Cs , with half lives of 30 yr and 2.06 yr respectively¹. For many years²⁻⁵ these radionuclides have been used as oceanographic tracers in the north-east Atlantic Ocean. Their sources are the fuel reprocessing plants in Western Europe. The radiocaesium released from Chernobyl, however, was deposited very inhomogeneously⁶, thus seriously perturbing its use as a tracer. So, there is a need to identify another tracer with the same sources but unperturbed by input from Chernobyl. Technetium-99 is not sedimented to any great degree and thus as a conservative tracer⁷ can be considered as a candidate. In a recent study² we found the same dilution factor for technetium as for radiocaesium when comparing concentrations in the North Sea with those in coastal waters around Greenland, supporting the idea of technetium as a substitute tracer for radiocaesium. Here we report a study of the possible contributions to ^{99}Tc in order to determine the Chernobyl input. We chose the Baltic Sea for our study because it probably shows higher concentrations of Chernobyl debris than any other sea. We show that the Chernobyl contribution is such that ^{99}Tc can be used as a tracer.

There are three possible main contributors to the ^{99}Tc content of the Baltic Sea; fallout from nuclear weapons testing in the 1950s and 1960s, liquid discharge from nuclear reprocessing in Western Europe and the Chernobyl accident. To test our

hypothesis seawater samples were collected during an FS *Gauss* cruise (arranged by the German Hydrographic Institute) to the Baltic Sea in October–November 1986, which were subsequently analysed for ^{99}Tc , ^{137}Cs and ^{134}Cs .

From seaweed measurements in earlier years⁸, we know that only very small amounts of ^{99}Tc have entered the Baltic and we would therefore expect a low background when measuring the Chernobyl contribution of ^{99}Tc . Thus, if Chernobyl-produced ^{99}Tc was indeed present in the marine environment, the Baltic Sea would be a likely place to detect it. Moreover, the Baltic is one of the seas most contaminated by Chernobyl fallout^{9,10}.

The samples analysed consisted of 200 l of sea water. After collecting, $^{99\text{m}}\text{Tc}$ yield tracer (metastable form of ^{99}Tc) was added to the water on board the ship. The tracer showed that ferrihydroxide precipitation under reducing conditions collected the technetium in nearly 100% yield. Technetium analysis was performed after the radiochemical solvent extraction from the hydroxide precipitate¹¹.

Gamma-ray spectra of the plated Tc isolates revealed radionuclides other than ^{99}Tc , principally ^{106}Ru , ^{103}Ru and $^{110\text{m}}\text{Ag}$. Thus the radiochemical procedure used hitherto could not be applied, and it was necessary to develop a new method¹² which could effectively remove these elements. In this method Tc and Ru are oxidized to TcO_4^- and RuO_4^- by NaClO and separated from other radionuclides by extraction with CCl_4 at $\text{pH} = 4$. The ruthenium is reduced to low valence, whereas Tc is maintained as TcO_4^- using H_2O_2 . Thus separated, Tc, Ru and the other radionuclides are subsequently isolated by solvent extraction with cyclohexanone and 5% triisooctylamine/xylene. Technetium is electroplated onto stainless steel disks. The radiochemical yield is determined by gamma-counting $^{99\text{m}}\text{Tc}$ prior to low-level beta-counting¹¹. The decontamination factor for this method is 10^5 for Ru and 2×10^5 for Ag. The lower limit of detection for a 200 l seawater sample is $0.014 \text{ Bq } ^{99}\text{Tc m}^{-3}$.

Figure 1 shows the distribution of ^{99}Tc in the Baltic Sea. Samples from the Cattegat contained about $1-2 \text{ Bq } ^{99}\text{Tc m}^{-3}$, as expected from earlier studies¹³. In the Baltic Sea the levels drop below 0.1 Bq m^{-3} . If the ^{99}Tc was from Chernobyl we would have expected to find the highest levels in the surface water, as observed for ^{137}Cs ; instead, the ^{99}Tc concentrations in the Baltic

Table 1 Technetium-99 and radiocaesium in Baltic seawater samples collected in 1986

Location		Date	Depth (m)	Salinity (‰)	¹³⁷ Cs	¹³⁴ Cs/ ¹³⁷ Cs	⁹⁹ Tc
N	E				(Bq m ⁻³)		(Bq m ⁻³)
57°00'	12°00'	15/10	4	23.4	106	0.35	2.10
57°00'	12°00'	15/10	40	32.7	84	0.35	1.15
56°45'	11°00'	15/10	4	22.6	100	0.32	1.26
56°30'	11°30'	15/10	4	22.6	96	0.39	1.09
56°30'	12°00'	16/10	4	21.9	97	0.36	1.23
55°17'	12°33'	16/10	4	8.22	33.4	0.24	—
55°34'	15°09'	17/10	4	7.76	37.1	0.35	0.040
55°34'	15°09'	17/10	~50	10.02	36	0.36	0.160
55°34'	15°09'	17/10	~70	13.62	43	0.30	0.160
57°00'	17°30'	17/10	4	6.82	265	0.51	0.017
57°00'	17°30'	17/10	~50	7.33	55	0.45	0.031
57°00'	17°30'	17/10	~90	9.46	18	0.11	0.028
57°00'	17°30'	17/10	4	6.82	259	0.52	0.055
58°53'	19°50'	18/10	4	6.43	753	0.51	0.016
58°53'	19°50'	18/10	50	7.06	132	0.53	0.025
58°53'	19°50'	18/10	130	10.13	18	0.10	0.061
58°45'	18°30'	18/10	4	6.73	449	0.42	0.043
58°45'	18°30'	18/10	120	9.93	16	0	0.072
58°45'	18°30'	18/10	250	10.24	19	0	0.049
60°13'	19°04'	21/10	4	5.51	962	0.51	0.047
60°13'	19°04'	21/10	120	7.25	152	0.47	0.055
60°13'	19°04'	21/10	240	7.45	70	0	0.041
61°30'	17°59'	22/10	4	5.50	592	0.50	0.082
62°40'	19°33'	23/10	4	5.78	523	0.52	0.038
62°40'	19°33'	23/10	~50	5.78	464	0.54	0.069
62°50'	19°33'	23/10	130	6.75	60	0.47	0.069
64°05'	21°56'	23/10	4	3.60	112	0.48	0.183
64°05'	21°56'	23/10	90	4.23	370	0.46	0.054
64°44'	23°06'	24/10	4	3.71	152	0.51	0.036
62°00'	20°30'	25/10	4	5.96	721	0.50	0.036
61°04'	19°42'	26/10	4	5.96	550	0.46	0.033
61°04'	19°42'	26/10	80	6.65	104	0.46	0.023
61°04'	19°42'	26/10	130	6.05	126	0.49	0.039
60°10'	26°35'	28/10	4	5.54	399	0.50	—
59°30'	23°20'	30/10	4	6.88	85.2	0.43	0.039
56°05'	17°42'	1/11	4	7.36	38.0	0.31	—
56°05'	17°42'	1/11	50	8.03	21	0.27	0.073
54°48'	19°19'	1/11	4	7.51	429	0.37	—
54°48'	19°19'	1/11	100	10.47	22	0.19	0.140

Most of the radiocaesium analyses were performed by the German Hydrographic Institute. Relative counting error for ⁹⁹Tc: 1–2 Bq m⁻³: 2–3%; 0.1–0.2 Bq m⁻³: 5–10%; <0.1 Bq m⁻³: 10–30%. The missing ⁹⁹Tc data were due to loss of sample during analysis.

Proper and the Bothnian Sea were generally higher below the halocline (at ~50 m) than above. This is in accordance with expectations if the principal source of ⁹⁹Tc in the Baltic Sea is liquid discharge from reprocessing operations in western Europe. This would be mixed with North Sea water of high salinity entering the Baltic Sea through the Danish Straits as a bottom current.

If the Chernobyl accident had contributed significant amounts of ⁹⁹Tc to the Baltic Sea, we would have expected a correlation between the amounts of radiocaesium and ⁹⁹Tc in the seawater samples. But even the samples with the highest ¹³⁷Cs concentrations (500–1,000 Bq m⁻³) showed no correlation between radiocaesium and ⁹⁹Tc. As the K_d value for Tc in coastal waters is 30 times lower⁷ than that of Cs it is unlikely that Tc should have sedimented relative to Cs and thus have given rise to an underestimate of Chernobyl-derived radiocaesium. The most northerly sample in the Bothnian Bay contained 0.18 Bq ⁹⁹Tc; this is more than was seen in the other Baltic surface samples. We can offer no explanation for this outlier.

The fission yield of ⁹⁹Tc (half life = 2.1×10^5 yr) (ref. 1) from thermal fission of ²³⁵U is 61,400 atoms per 10^6 fissions (ref. 14). This is similar to the production of ¹³⁷Cs (half life = 30 yr), at 59,400 atoms per 10^6 fissions. Hence the theoretical activity ratio ⁹⁹Tc/¹³⁷Cs in reactor debris should be $\sim 1.5 \times 10^{-4}$. If this ratio had been maintained in Chernobyl debris, we would have expected seawater concentrations of ⁹⁹Tc in the most contaminated

samples to have been a factor of two to three higher than the levels observed.

Due to similarities in the chemical properties and volatility of Tc and Ru (ref. 15), it is likely that the release percentage of Tc from Chernobyl was similar to that of Ru. According to USSR information¹⁶, 2.9% of the core inventory of Ru was released. For ¹³⁷Cs the percentage was quoted as 13%, but several investigators^{17,18} have found that this release must be too low by a factor of two to three. Therefore, we suggest that the release percentage of ⁹⁹Tc was an order of magnitude less than that of ¹³⁷Cs, in which case the ⁹⁹Tc/¹³⁷Cs ratio in Chernobyl debris should be 1.5×10^{-5} . The inventory of Chernobyl ¹³⁷Cs is estimated from a mean deposit of 10 kBq ¹³⁷Cs m⁻² (ref. 6) which, in the area of the Baltic Sea (374,000 km²), gives a total ¹³⁷Cs deposition of 3.7 PBq. The estimated ⁹⁹Tc deposition from Chernobyl then becomes 0.05 TBq.

To measure ⁹⁹Tc in the Chernobyl debris directly, we analysed a highly contaminated sample. This was a motor-filter from a fishing boat, which had passed through the Chernobyl cloud on 27–28 April 1986 in the western Baltic Sea at Bornholm. This sample showed a ⁹⁹Tc/¹³⁷Cs ratio of $(1.01 \pm 0.13) \times 10^{-5}$ (± 1 standard error from four measurements). The measured ratio was thus in reasonable agreement with the above estimate. We also analysed a grass-turf collected in Kiev in September 1986, which contained no detectable amount of ⁹⁹Tc. Our detection

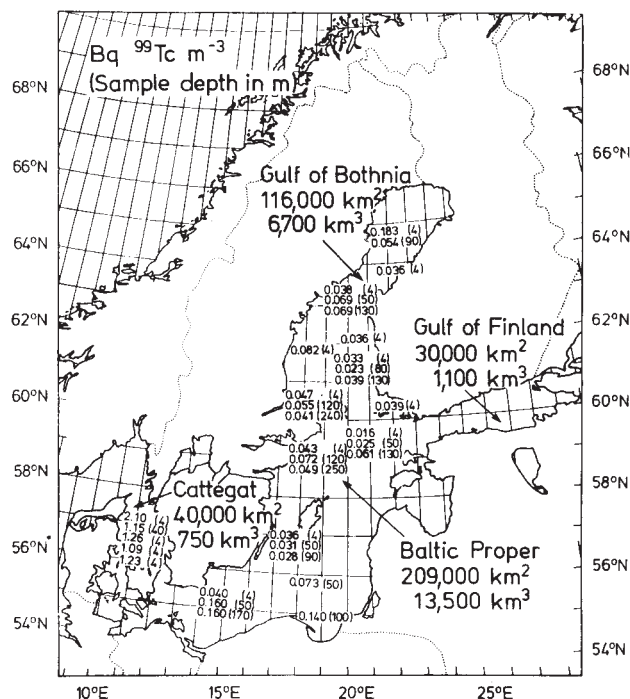


Fig. 1 Technetium concentrations in Baltic seawater collected in October–November 1986.

limit implies that $^{99}\text{Tc}/^{137}\text{Cs}$ was less than 5×10^{-4} , again consistent with the above.

Salo *et al.*¹⁹ have calculated the global fallout inventories of ^{90}Sr and ^{137}Cs in the Baltic Sea. The calculated inventory in any year was obtained²⁰ by adding the contributions from direct deposition on the sea surface (precipitation), run-off (with rivers) and inflow through the Danish Straits, and subtracting the outflow through the Straits and the radioactive decay. If the subtraction of decay is omitted, we obtain the inventory that would pertain if the radionuclides (^{90}Sr or ^{137}Cs) had been stable and thus comparable with ^{99}Tc , which in practice behaved as a stable isotope during the period of observation. As the sedimentation and runoff characteristics of ^{99}Tc are closer to those of ^{90}Sr than to ^{137}Cs (refs 2, 21), we used the ^{90}Sr data. In 1981 Salo *et al.* estimated a ^{90}Sr inventory of 552 TBq; this corresponds to an undecayed amount of 811 TBq ^{90}Sr . According to UNSCEAR²², the $^{137}\text{Cs}/^{90}\text{Sr}$ ratio in global fallout is 1.6; therefore, the $^{99}\text{Tc}/^{90}\text{Sr}$ ratio is 1.6 times the $^{99}\text{Tc}/^{137}\text{Cs}$ ratio in undecayed global fallout. Hence we estimate the inventory of global fallout-derived ^{99}Tc in the Baltic as 0.2 TBq.

The total inventory of ^{99}Tc in the Baltic Sea was calculated as follows. In the Gulf of Bothnia (comprising the Bothnian Bay and the Bothnian Sea) we assumed no halocline. Thus we used the mean of all eleven measurements, $0.060 \text{ Bq } ^{99}\text{Tc m}^{-3}$, multiplied by the water volume, $6,700 \text{ km}^3$ (ref. 19) giving an inventory of $0.40 \text{ TBq } ^{99}\text{Tc}$. In the Baltic Proper (see Fig. 1) and the Gulf of Finland there is a halocline at $\sim 50 \text{ m}$ depth. The water volumes above and below the halocline are $11,100 \text{ km}^3$ and $3,500 \text{ km}^3$ respectively. From Table 1 we calculated the mean of the eleven samples from the Baltic Proper with a sampling depth $\leq 50 \text{ m}$ and of the eight samples collected below 50 m . These means, 0.050 and $0.076 \text{ Bq } ^{99}\text{Tc m}^{-3}$ respectively, were multiplied by their respective volumes, giving $0.55 \text{ TBq } ^{99}\text{Tc}$ above the halocline, 0.27 TBq below and a total inventory of $1.22 \text{ TBq } ^{99}\text{Tc}$. The latter is similar to the mean of all 30 samples (excluding those from the Cattegat) multiplied by the total volume, which gives $1.28 \text{ TBq } ^{99}\text{Tc}$. Our best estimate of the inventory is $1.3 \text{ TBq } ^{99}\text{Tc}$, and thus our estimate of the contribution from reprocessing is $1.05 \text{ TBq } ^{99}\text{Tc}$. As it takes 5–6 yr for discharges from Sellafield to enter the Baltic Proper¹⁹, this

amount of ^{99}Tc represents discharges up to 1980. Compared with the $0.6 \text{ PBq } ^{99}\text{Tc}$ discharged from Sellafield from 1970 to 1980²³, the ^{99}Tc inventory in the Baltic Sea from reprocessing amounts to $\sim 0.18\%$ of the total release.

As the Baltic was the sea where maximum water concentration of ^{99}Tc was most likely to occur after the Chernobyl accident, and as we were unable to detect any additional ^{99}Tc in the Baltic sea water after the accident, we conclude that the Chernobyl accident did not contribute significantly to ^{99}Tc levels in the north-eastern Atlantic Ocean. Discharges from nuclear reprocessing in western Europe remain the dominant source of this radionuclide.

This project was supported by the CEC Radiation Protection Programme. Dr Chen had been awarded a fellowship under the International Atomic Energy Agency. We thank the F/S *Gauss* and its crew for hosting us during the cruise to the Baltic Sea.

Received 7 March; accepted 3 August 1988.

1. ICRP Publication 38 (Pergamon, Oxford–New York–Frankfurt, 1983).
2. Aarkrog, A. *et al. Estuar. Coast. Shelf Sci.* **24**, 637–647 (1987).
3. Jefferies, D. F., Steele, A. K. & Preston, A. *Deep Sea Res.* **29**, 713–738 (1982).
4. Kautsky, H. *Di. hydrogr. Z.* **40**, 49–69 (1987).
5. Livingston, H. D., Bowen, V. T. & Kupferman, S. L. *J. mar. Res.* **40**, 1227–1258 (1982).
6. Devell, L., Aarkrog, A., Blomqvist, L., Magnusson, S. & Tveten, U. *Nuclear Europe VI*, 16–17 (1986).
7. *Tech. Rep. Ser. 247: Sediment K_d s and Concentration Factors for Radionuclides in the Marine Environment* (Int. Atomic Energy Agency, Vienna, 1985).
8. Holm, E., Riosco, J. & Mattsson, S. in *Technetium in the Environment* (eds Desmet, G. & Myttenaere, C.) 61–68 (Elsevier, 1984).
9. *Meereskundliche Beobachtungen und Ergebnisse Nr 62* (Deutsches Hydrographisches Institut, Hamburg, 1987).
10. Buesseler, K. O. *et al. Nature* **329**, 825–828 (1987).
11. Holm, E., Riosco, J. & Garcia-Leon, M. *Nucl. Instrum. Meth. Phys. Res.* **223**, 204–207 (1984).
12. Chen, Q., Aarkrog, A., Dick, H. & Mandrup, K. *Risø-M-2671* (Risø National Laboratory, 1987).
13. Aarkrog, A. *et al. Risø-R-540* (Risø National Laboratory, 1987).
14. Harley, J. H. *HASL-300* (Environmental Measurements Laboratory, New York, 1972).
15. Till, J. E. in *Technetium in the Environment* (eds Desmet, G. & Myttenaere, C.) 1–20 (Elsevier, 1984).
16. *Summary Report on the Post-Accident Review Meeting on the Chernobyl Accident Safety Series No 75-INSAG-1* (IAEA, Vienna, 1986).
17. ApSimon, H. M., Macdonald, H. F. & Wilson, J. J. N. *J. Soc. radiol. Prot.* **6**, 109–119 (1986).
18. *Environmental Health Series No 24: Health Hazards from Radiocaesium Following the Chernobyl Nuclear Accident* (World Health Organization, Copenhagen, 1987).
19. *Study of Radioactive Materials in the Baltic Sea* IAEA-TEC-DOC-362 (Int. Atomic Energy Agency, Vienna, 1986).
20. Salo, A. & Voipio, A. *Mar. Pollution Bull.* **12**, 218–224 (1981).
21. Blaylock, B. G., Frank, M. L., Hoffman, F. O. & De Angelis, D. L. in *Technetium in the Environment* (eds Desmet, G. & Myttenaere, C.) 79–89 (Elsevier, 1984).
22. *UNSCEAR, Ionizing Radiation: Sources and Biological Effects* (United Nations, New York, 1982).
23. *BNFL Annual Reports on Radioactive Discharges and Monitoring of the Environment* (British Nucl. Fuels Ltd, Risley, 1978–85).

Seismological and structural evolution of strike-slip faults

Steven G. Wesnousky

Center for Earthquake Research and Information,
Memphis State University, Memphis, Tennessee 38152, USA

The mapped traces of strike-slip faults are commonly characterized by discontinuities that appear as steps in map-view. Here I present observations to show that the number of steps per unit length along the trace of major strike-slip fault zones in California and Turkey is a smoothly decreasing function of cumulative geological offset. When coupled with a growing body of evidence that indicates that steps in fault traces work to impede or arrest the propagation of earthquake ruptures, the apparent smoothing of fault traces with displacement is interpreted to suggest that the spatial distribution of strength properties on a fault plane is a function of cumulative geological offset. A consequence of this structural evolution is that faults may undergo a seismological evolution a well, whereby the size and frequency distribution of earthquakes is also a function of cumulative offset.

An earthquake in California or other region of similar tectonic style does not generally rupture the whole length of the fault

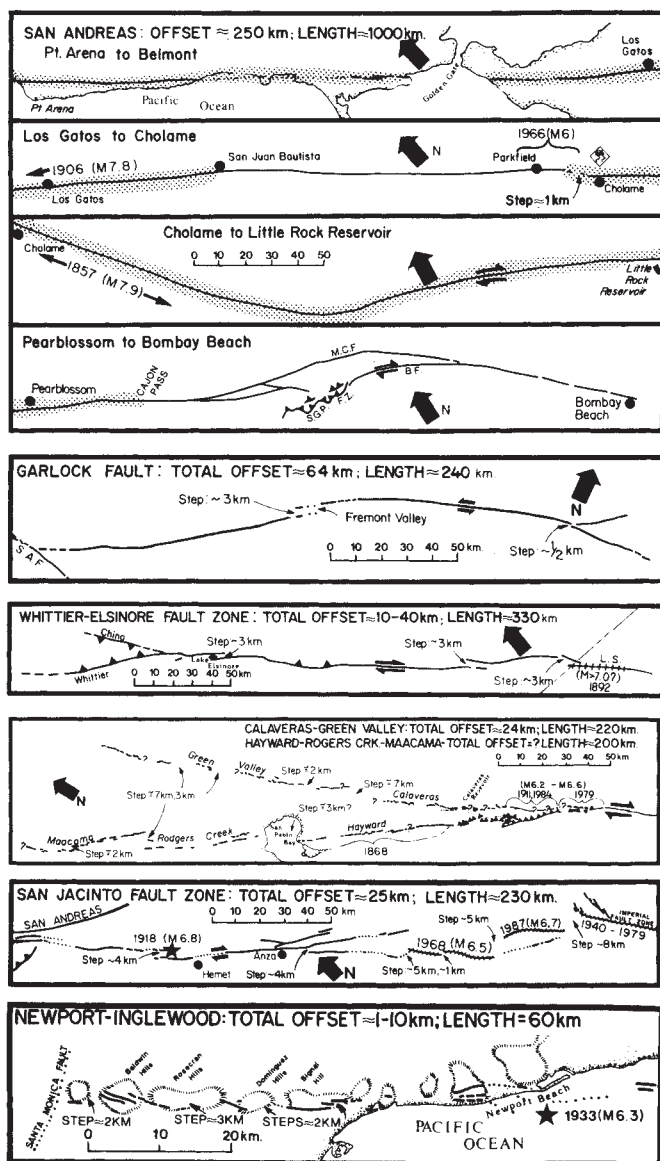


Fig. 1 Maps of the San Andreas¹⁶⁻²², Garlock²³, Whittier-Elsinore²⁴, Calaveras-Green Valley²⁵, San Jacinto²⁶ and Newport-Inglewood¹³ fault zones of California showing the location of steps in fault traces which are characterized by stepover widths $W_s \geq 1$ km. Segments of fault which have ruptured during historical earthquakes are marked by hachures, stippling or brackets, and the dates and magnitudes of the respective earthquakes. Epicentre of 1933 Long Beach earthquake is marked by star and half-sided arrows indicate sense of displacement along faults.

on which it occurs but, rather, only a fraction of the entire fault length¹. Observations also show that the endpoints of strike-slip earthquake ruptures are commonly associated with step-like discontinuities along mapped fault traces^{2,3}. When slip occurs along a fault during an earthquake, stress concentrations will occur at step-like discontinuities along the fault². Such discontinuities are commonly referred to as releasing- or restraining-steps, depending on whether the stress concentration is extensional or compressional in nature⁴, respectively, and the severity of the stress concentration will generally be a function of the width W_s of the step as measured perpendicular to fault strike². The stress concentrations at steps may be sufficient to impede or stop the propagation of an earthquake rupture and, on that basis, a number of investigators have attributed a causal association to the observed coincidence between the endpoints of earthquake ruptures and the location of steps along mapped fault traces^{3,5}. In turn, laboratory models and scattered field

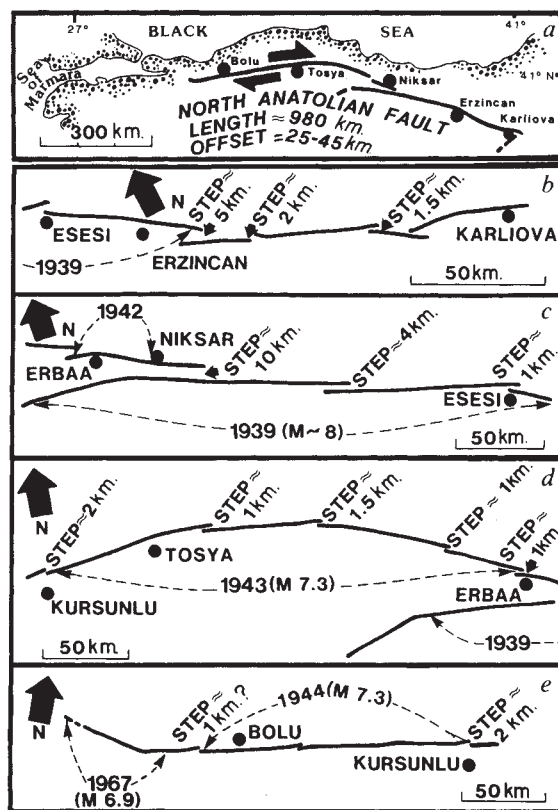


Fig. 2 a, The North Anatolian fault shows right-lateral movement and strikes westward across Turkey from a point near Karliova to west of Bolu. Larger-scale and overlapping strip maps of the fault (b-e) are taken from Barka and Kadinsky-Case⁵ and show location of steps greater than 0.1 km in size. Dates and arrows show the year and extent of rupture during large historical earthquakes, respectively.

observations have been the basis to suggest that, during the initial stages of development, strike-slip faults are characterized by a zone of discrete fault segments, each of which is separated from the next segment by a step, and that, with continued displacement, the individual fault segments ultimately coalesce to accommodate displacements along a principal and through-going fault^{6,7}. The few observations available thus point to a close connection between the seismological and the structural evolution of fault zones and provided the impetus for this work: to examine whether or not there exists a systematic relationship between frequency of occurrence of step-like discontinuities and the total geological offset registered along the trace of active strike-slip faults.

Strike-slip fault zones and earthquakes in California and Turkey are the principal basis for my observations. It is for these regions that the most comprehensive maps and data concerning the total geological offset and the location of historical earthquakes along the trace of active faults are available. More specifically, I have relied on the published literature to obtain estimates of the cumulative geological offset across the San Andreas⁸, Garlock⁹, Whittier-Elsinore¹⁰, Calaveras-Green Valley¹¹, San Jacinto¹² and Newport-Inglewood¹³ faults of California and the North Anatolian fault zone¹⁴ of Turkey. Similarly, the magnitude and rupture length of historical ruptures along strike of each of the fault zones has been the source of numerous previous studies^{1,5}. Thus, the location and magnitude of historical earthquake ruptures, the total fault length, the cumulative geological offset and the locations and sizes of step-like discontinuities along strike of each of the fault zones considered have been detailed in a separate report¹⁵ and are summarized in the accompanying maps (Figs 1 and 2). Because of the scale of

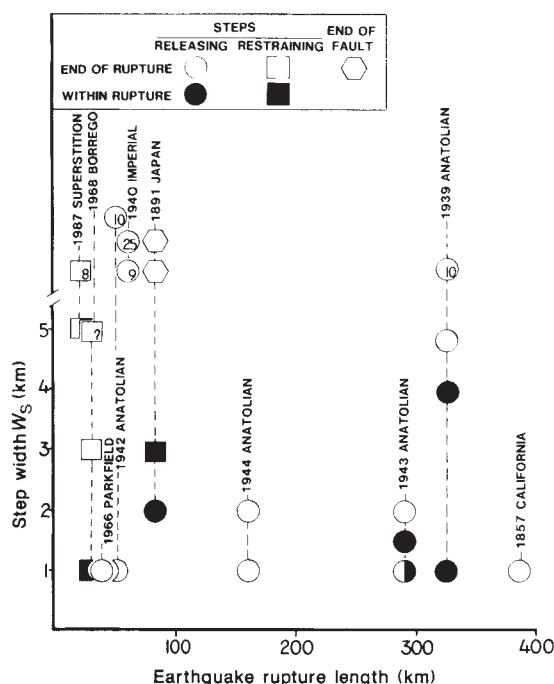


Fig. 3 Earthquake rupture length versus the size W_s of steps in mapped fault trace observed at the endpoints (open symbols) and within (solid symbols) the rupture zone of large Anatolian and California earthquakes. Half-filled symbol indicates that steps of similar size occurred both within and at endpoint of rupture. Steps in fault trace are also subdivided according to whether they are restraining (squares) or releasing (circles) in nature.

available maps, I have limited attention to steps in fault trace characterized by stepover widths of a kilometre or greater.

Figure 2 illustrates the relationship of the size and location of steps along the traces of strike-slip faults to the length and location of earthquake ruptures. The figure is a plot of the rupture lengths of large California and Anatolian earthquakes versus the stepover width W_s for each step in the fault trace observed at both the endpoints and within the respective earthquake rupture zones. For example, Fig. 3 shows that the 1939 Anatolian earthquake produced ~320 km of surface rupture and that the endpoints of the rupture are associated with steps of about 5 and 10 km respectively. The figure also shows that the same rupture propagated through steps of 1 and 4 km; that is to say, the steps located within the 1939 rupture zone are smaller than the steps associated with the termination of the rupture. A similar pattern is found for the other earthquakes in Fig. 3, that is, steps associated with the endpoints of ruptures are commonly the largest discontinuities registered along the length of the respective rupture zones. The trend, admittedly, is not without exception. The 1943 Anatolian earthquake is a case in point, where a step interior to the rupture zone is larger than a similar step observed at the endpoint of the rupture. It is also apparent that the endpoints of all ruptures are not marked by steps in mapped fault traces, as evidenced by the southern limit of the 1857 California earthquake (Fig. 1). Nonetheless, the observations are certainly sufficient to argue that the association of steps with the endpoints of earthquakes is not merely coincidence. Rather, the stress concentrations resulting from the discontinuities appear to play a principal and causal role in limiting the extent of earthquake ruptures.

A number of recent studies have attempted to explain characteristics of the spatial, temporal and size distributions of earthquakes along faults by assuming an arbitrary or fractal distribution of strength properties along fault planes²⁷⁻³¹. The observations I have summarized thus far suggest that the

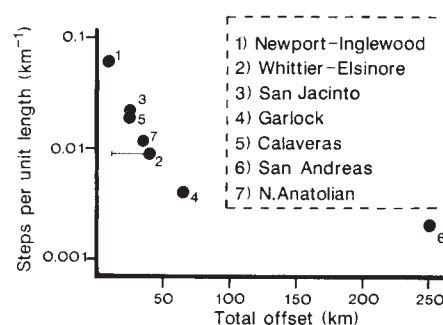


Fig. 4 The number of steps ($W_s \geq 1$ km) per unit length of mapped fault trace versus cumulative geological offset along major strike-slip faults in California and Turkey.

geometrical complexity of fault traces may be used to gauge, in a general sense, the heterogeneity of strength properties along fault planes. Toward that end, the number of steps ($W_s \geq 1$ km) per unit length of a fault zone provides a simple, quantitative measure of fault trace complexity. The complexity of fault trace measured in this manner is plotted as a function of cumulative geological offset in Fig. 4. Fault trace complexity varies by nearly two orders of magnitude within the data set and is a smoothly decreasing function of cumulative geological offset. The observation supports the idea that fault zones are characterized by a discontinuous trace during the initial stages of development and that increasing displacement along such faults works to remove discontinuities along the fault trace, effectively smoothing the fault plane. Moreover, when coupled with the growing body of evidence that indicates that steps in fault traces work to impede or arrest the propagation of earthquake ruptures, the apparent smoothing of the fault trace with displacement may be further interpreted to indicate that the distribution of strength properties on a fault plane is spatially heterogeneous, and that the degree of heterogeneity is a function of cumulative offset. A consequence of this structural evolution is that faults may undergo a seismological evolution as well, whereby the size and frequency distribution of earthquakes on a fault is also a function of cumulative geological offset.

I thank A. Barka and A. Johnston for reviews of the manuscript, A. Barka, T. Rockwell and D. Herd for preprints of unpublished papers and maps, Tanya George for drafting the figures and the NSF for supporting the research.

Received 26 May; accepted 12 August 1988.

- Wesnowsky, S. G. *J. geophys. Res.* **91**, 12587-12631 (1986).
- Segall, P. & Pollard, D. D. *J. geophys. Res.* **85**, 4337-4350 (1980).
- Sibson, R. H. *Geophys. Monogr.* **37**, 157-167 (Am. Geophys. Un., Washington D.C., 1986).
- Crowell, J. C. *Soc. econ. Paleontologists and Mineralogists Spec. Pap.* **22**, 190-204 (1974).
- Barka, A. A. & Kadinsky-Cade, C. *Tectonics* (in the press).
- Wilcox, R. E., Harding, T. P. & Seely, D. R. *Bull. Am. Ass. Petrol. Geol.* **57**, 74-94 (1973).
- Withjack, M. O. & Jamison, W. R. *Tectonophysics* **126**, 99-124 (1986).
- Hill, M. L. *Geol. Soc. Am. Bull.* **92**, 112-131 (1981).
- Smith, G. I. *Bull. Am. Assoc. Petrol. Geol.* **46**, 185-104 (1962).
- Lamar, D. L. & Rockwell, T. K. in *Neotectonics and Faulting in Southern California, Guidebook and Volume 82nd Annual Meeting of the Cordilleran Section of the Geol. Soc. Am.* 149-158 (Los Angeles, California, 1986).
- Kintzer, F. C., Brooks, E. R. & Cummings, J. C. *Geol. Soc. Am. Abstracts with Programs* **9**, 65 (1977).
- Sharp, R. V. *Geol. Soc. Am. Bull.* **78**, 705-730 (1967).
- Barrows, A. G. *Spec. Rep. Calif. Div. Mines Geol.* **114**, 115 (1974).
- Barka, A. A. & Gulen, L. *Spec. Publ. Middle East. Univ.* (in the press).
- Wesnowsky, S. G. *Proc. USGS Workshop on Fault Segmentation and Controls of Rupture Initiation and Termination - USGS Open-File Report* (in the press).
- Clark, M. M. *USGS Misc. Field Invest. Map* **1-1483** (1984).
- Hope, R. A. *USGS Open-file Rep.* **69-130** (1969).
- Ross, D. C. *USGS Misc. Field Invest. Map* **1-553**, (1969).
- Vedder, J. G. & Wallace, R. E. *USGS Misc. geol. Invest. Map* **1-574** (1970).
- Brown, R. D. *USGS Misc. geol. Invest. Map* **1-575** (1970).
- Brown, R. D. Jr & Wolfe, E. W. *USGS Misc. geol. Invest. Map* **1-692** (1972).
- Matti, J. C., Morton, D. M. & Cox, B. F. *USGS Open-file Rep.* **85-365**, 23 (1985).
- Clark, M. M. *USGS Misc. geol. Invest. Map* **74**, (1973).
- Anderson, J. G., Rockwell, T. & Agnew, D. *Earthquake Spectra* (in the press).
- Herd, D. G. *USGS Misc. Field Stud.* **1:250,000 Map of Recently Active Faults of San Francisco Bay Region and Surrounding Northern California Coast Ranges (in the press).**
- Sharp, R. V. *Bull. Calif. Div. Mines Geol.* **196**, 187-194 (1975).

27. Lay, T. & Kanamori, H. in *Earthquake Prediction: an International Review* (eds Simpson, D. & Richards, P.) 579-592 (Maurice Ewing Series 4, Am. Geophys. Un., Washington, D.C., 1981).
28. Andrews, D. J. *J. geophys. Res.* **85**, 3867-3877 (1980).
29. Aki, K. in *Earthquake Prediction - An International Review* (eds Simpson, D. & Richards, P.) 566-574 (Maurice Ewing Series 4, A. Geophys. Un., Washington, D.C., 1981).
30. Scholz, C. H. & Aviles, C. A. in *Earthquakes Source Mechanics* Geophys. Monogr. 37, 147-156 (eds Das, S., Boatwright, J. & Scholz, C.) (Maurice Ewing Series 6, Am. Geophys. Un., Washington, D.C., 1986).
31. Smalley, R., Turcotte, D. & Solla, S. *J. geophys. Res.* **90**, 1894-1900 (1985).

An experimental determination of primary carbonatite magma composition

Margaret E. Wallace & David H. Green

Geology Department, University of Tasmania, GPO Box 252C, Hobart, Australia 7001

Carbonatites are uncommon carbonate-rich rocks usually found in continental intra-plate regions and often associated with rifting. There has been much debate as to whether carbonatite magmas are primary melts derived from partial melting of mantle peridotite, or are formed by exsolution of an immiscible carbonate melt fraction from phonolitic or nephelinitic magmas. Our experiments on the phase relationships of carbonate and amphibole-bearing peridotite (containing 0.3% H₂O and 0.5-2.5% CO₂) show that sodic dolomitic carbonatite magma coexists with an amphibole lherzolite assemblage in a field ranging from 21 to 30 kbar and 930 to 1,080 °C, spanning a pressure and temperature interval between the solidus and the amphibole breakdown and melting curve. Thus primary carbonatite melts may occur under suitable geothermal conditions. The nature of the peridotite solidus and of the melting reactions differ considerably from published models¹⁻³. The carbonatite melt composition, determined by a series of 'sandwich' experiments, was found to be rich in Na, Mg, Ca and Fe, with a small dissolved silicate content. This melt quenches to an assemblage of dolomite and Na-Mg carbonate minerals, producing textures similar to those preserved in samples from Oldoinyo Lengai⁴, Homa mountains, Tanzania⁵ and Kaiserstuhl, Germany⁶.

The majority of carbonites are rich in calcium and are composed of calcite (the calcite carbonatites, sövite and alvikite) and dolomite (beforsites or dolomite carbonatites) with minor amounts of ferroan carbonate. Cation abundance is in the order Ca > Mg > Fe > (Na + K) (ref. 7). These minerals characteristically contain high contents of Sr, Ba, Nb and rare earth elements, and are associated with ijolites and other alkali-rich igneous rocks, and commonly are surrounded by large sodic or potassic fenitization halos. A particularly important type of carbonatite has been erupted by Oldoinyo Lengai, northern Tanzania⁴. It is highly alkalic (Table 1, column 12) and is composed dominantly of Na₂O, K₂O, CaO and CO₂, which form the mixed carbonate minerals nyerereite (Na_{0.22}K_{0.18})₂Ca(CO₃)₂ and gregoryite (Na_{0.78}K_{0.05})₂Ca_{0.17}CO₃ (ref. 8). Several other carbonatite lavas, now composed largely of calcite, have been interpreted recently as originally having had similar mineralogy⁹, subsequently replaced by reaction with calcium-rich ground water^{4,5,9}. It is thus argued that alkali-rich carbonatite magma occurred more frequently than is indicated by the present relative abundance of calcitic carbonatites and natro-carbonatites.

Isotope studies demonstrate that carbonatitic magmas are derived from mantle source areas¹⁰⁻¹², but do not distinguish between origins by partial melting of the mantle or by exsolution from a phonolitic or nephelinitic magma^{7,13-18}. Experimental investigations of carbonatite genesis have explored the CaO-MgO-SiO₂-CO₂-H₂O system, its subsystems and other systems involving alkali elements with CO₂ and H₂O^{14,19-21}. This extensive literature has been summarized by Wyllie²² and Eggler²³.

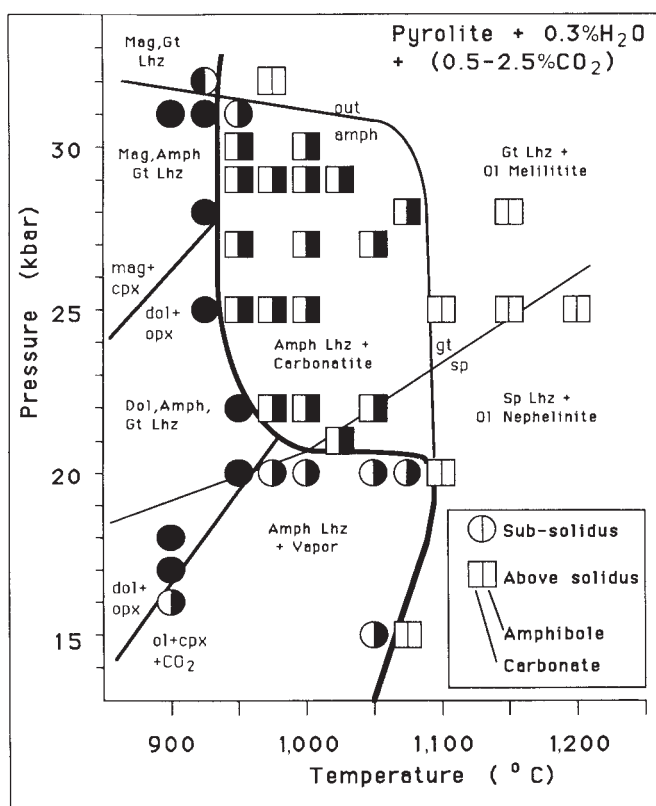


Fig. 1 Phase relationships of amphibole-rich Hawaiian Pyrolite (0.3% H₂O) plus 5% dolomite or 1.4% magnesite. Carbonatite melt field is bounded by the carbonate solidus at 930 °C (heavy line) and amphibole breakdown and silicate melt field above 1,080 °C (25 kbar). Size of the box symbols indicates the uncertainty in pressure and temperature. All runs were buffered with magnesite ($f_{O_2} \approx \text{EMOG}$).

These earlier studies confirmed the existence of low-temperature alkali-rich carbonate melts, explored liquid immiscibility between silicate and carbonate-rich melts, and documented stability fields for carbonated peridotite under upper-mantle conditions. Other studies have investigated the solubility of CO₂ as (CO₃²⁻) in silicate melts at high pressure²⁴⁻²⁵ and have emphasized continuity at high pressures between kimberlite, olivine melilitite and carbonatite magma compositions. Although there is thus a generally accepted framework for carbonatite genesis in the upper mantle, there is a continuing debate on the relative importance of crystal fractionation (for example, to yield Oldoinyo Lengai natrocarbonatite magma as a differentiate from parental calcitic or dolomitic carbonatite¹⁷) and liquid immiscibility (for example, to yield Oldoinyo Lengai natrocarbonatite magma as an exsolution from parental carbonated nephelinite or phonolite magma¹⁸). The specific problem of accurate determination of both the solidus temperature of a carbonate-bearing lherzolite and the nature of the melt at the solidus at high pressures has frustrated experimental petrologists and generated much debate^{1-3,22-27}. We have clarified the position of the peridotite solidus and we demonstrate that, because of the importance of the minor components Na, K, P and Ti, primary mantle melts of sodic, dolomitic carbonatite character can be produced under oxidizing conditions from amphibole and carbonate-bearing lherzolite.

Phase relationships for (amphibole pyrolite minus 40% olivine²⁷ plus 1.4% MgCO₃ or 5% CaMgCO₃) are presented in Fig. 1. The (amphibole pyrolite minus 40% olivine) starting-mix was prepared in 100-mg batches²⁷ and all phases analysed by microprobe. The peridotite composition used represents a vapour-undersaturated 'fertile' mantle peridotite, in which par-

Table 1 Starting materials and run products, sample T-2518 (22 kbar, 1,000 °C, 3 h)

	1	2	3	4	5	6	7	8	9	10	11	12	13
P ₂ O ₅	0.06	—	0.48	—	—	—	—	—	—	—	4.22	1.13	3.04
SiO ₂	45.56	—	2.94	40.91	55.43	51.25	45.67	—	—	0.37	0.19	<0.10	—
TiO ₂	1.14	—	0.45	—	0.26	1.10	1.57	55.17	0.61	—	0.23	0.01	—
Al ₂ O ₃	6.02	—	1.95	—	2.71	5.18	11.69	0.32	58.53	0.28	—	0.03	0.13
Cr ₂ O ₃	0.62	—	0.22	—	0.50	0.95	1.40	0.87	8.02	—	—	—	—
FeO	7.44	5.20	4.61	12.32	7.90	4.04	4.86	29.75	12.27	4.35	9.78	0.26	2.71
NiO	0.12	—	—	0.45	—	—	—	—	0.35	—	—	—	—
MnO	0.12	—	—	—	—	—	—	—	—	—	0.25	—	—
MgO	29.49	16.70	14.19	46.03	32.25	17.50	18.74	13.89	20.04	16.70	8.88	0.26	0.26
CaO	6.41	19.80	21.29	0.30	0.94	19.24	11.00	—	0.18	30.50	8.75	13.93	16.82
K ₂ O	0.15	—	0.35	—	—	—	0.60	—	—	—	1.34	6.55	2.21
Na ₂ O	0.82	12.50	4.99	—	—	0.74	2.48	—	—	—	22.48	30.30	31.57
Cl	—	—	—	—	—	—	—	—	—	—	0.19	2.45	—
CO ₂	2.29	45.80	—	—	—	—	—	—	—	—	—	33.90	—
Totals	100.24	100.00	51.49	100.00	100.00	100.00	98.00	100.00	100.00	52.71	56.43	89.17	52.75
Mg*	87.6	85.1	84.5	86.9	87.9	88.5	87.3	45.4	74.4	87.2	61.8	—	—

1. Amphibole-pyrolite-40% olivine (0.3% H₂O)+5% dolomite used as peridotite layer, (90%) of starting mix.

2. Mixed carbonate composition, 10% of starting mix, added to charge as layer below column 1 peridotite layer.

3. Carbonatite melt produced in equilibrium with peridotite assemblage.

4-9. Silicate phases, 4 = olivine, 5 = orthopyroxene, 6 = clinopyroxene, 7 = amphibole, 8 = ilmenite, 9 = spinel.

10, 11. Quench minerals in carbonatite melt (3), 10 = dolomite, 11 = acicular, Na-carbonate.

12. Natro-carbonatite BD118 from Oldoinyo Lengai⁴ (rock also containing SO₃ = 3.23, F = 1.20, H₂O⁺ = 4.51).

13. Experimental carbonatite melt (3) minus 61% dolomite, 7% olivine, 3% spinel, 1% ilmenite.

Mg* is defined as 100 × Mg/(Mg + Fe).

gasitic amphibole, dolomite or magnesite are minor sub-solidus phases and are important in determining melting relationships. Experiments used a piston-cylinder apparatus and well established techniques²⁶⁻²⁷. Samples were placed in small capacity (2.2-mm outside diameter) Ag₇₅Pd₂₅ capsules, surrounded by magnesite containing a small water content and enclosed in large-capacity (3.5-mm outside diameter) Ag₅₀Pd₅₀ capsules. NaCl or NaCl-pyrex sleeves were used. Experiments were maintained in the carbonate stability field by the buffering reaction MgCO₃ → MgO + C + O₂. This produces values of *f*_{O₂} close to the EMOG (enstatite + magnesite + olivine + graphite) buffer²⁸. Analyses of phases and of area scans show that there was no detectable iron loss from the sample to the Ag₇₅Pd₂₅ container, provided suitable run times were chosen from between 3 h for above-solidus runs, to 28 h for sub-solidus runs. Pressures are accurate to within 1 kbar. Temperatures were measured using a Pt/Pt₉₀Rh₁₀ thermocouple and are accurate to within 10 °C.

Olivine (ol), orthopyroxene (opx), clinopyroxene (cpx) and garnet (gt) or spinel (sp) are present in all the run products. Additional phases include amphibole (amph), dolomite (dol), magnesite (mag), ilmenite, apatite and rutile. We have established the presence of a carbonatite melt field between 21 kbar and 31 kbar, bounded by a low-temperature melting reaction with loss of subsolidus dolomite or magnesite (the solidus for carbonatite melt) and the higher-temperature amphibole breakdown and melting reaction, above which liquids are CO₃²⁻-bearing silicate melts. At lower pressures (<21 kbar), instability of carbonate in the presence of pyroxene produces a higher-temperature solidus at the amphibole melting reaction in the presence of CO₂-rich fluid. At higher pressures (>31 kbar), amphibole instability produces a low-temperature vapour-saturated (H₂O-rich) solidus for magnesite-bearing lherzolite. Our observations on experimental charges at >31 kbar suggest that immiscible silicate (olivine melilititic to kimberlitic) and carbonatite melts may coexist at *T* ~ 1,000 °C, but become a single-phase CO₂-rich silicate melt at higher temperatures (*T* > 1,100 °C) because of high CO₃²⁻ solubility in silicate melts at these pressures²²⁻²⁵.

The shape of the carbonate solidus—concave towards higher temperatures at high pressures and concave towards lower tem-

peratures in the low-pressure region—is similar to experiments in the CaO-MgO-Al₂O₃-SiO₂ system²⁴. The high- and low-pressure regions are divided by a solidus inflexion at 21 kbar which is controlled by the reaction opx + dol → cpx + ol + vapour^{2,3,21-27}. This produces a sub-solidus assemblage of (amphibole, spinel lherzolite, sp lhz) + vapor at low pressures and (amphibole + dolomite (or magnesite) + garnet lherzolite, gt lhz) at high pressure. Above the breakdown of amphibole a marked increase in degree of melting and change in melt composition occurs by the reaction opx + amph + sp → ol + cpx + melt, which produces a nepheline-normative melt^{29,30}. At pressures >21 kbar the solidus is defined by the breakdown of dolomite and the sodium-rich jadeitic component of clinopyroxene to produce an aluminous phase plus melt. This reaction, jadeite_{ss} + olivine + dolomite → garnet + diopside + orthopyroxene + carbonatite melt, produces a melt at 930 °C, ~50 °C lower than can be produced from water-saturated peridotite without carbonate and ~220 °C lower than the dehydration solidus for vapour-absent amphibole pyrolite without carbonate in the 15–28-kbar pressure interval²⁷.

The temperature and pressure region in which the carbonate melt is stable depends on the stability of amphibole. This is a function of bulk rock composition and the activity of water. Amphibole is stable to the highest pressures and temperatures in a fertile peridotite where the mole fraction of water in the fluid species (*x*_{H₂O}) ≈ 0.4 (refs 27, 29). The pressure-temperature field of carbonatite melt and amphibole lherzolite is more limited for a refractory (Na-, Ti-, K-poor) lherzolite because of the lower temperature of amphibole breakdown³¹. Phase relations for carbonate-bearing, water-saturated amphibole lherzolite remain to be explored, but will lower the temperature of appearance of water-rich silicate melts from *T* > 1,080 °C to temperatures close to 1,000 °C in the pyrolite composition²⁷. The melt composition changes from carbonatite to olivine melilitite as amphibole breaks down or melts and carbonate dissolves in increasing amounts of silicate melt²⁵.

Our first intimation that the melt phase was strongly enriched in Na₂CO₃ was the observation of a small vein of Na₂CO₃-rich quench in an experiment on (amphibole-bearing pyrolite minus 40% olivine plus 5% dolomite) at 22 kbar, 975 °C. (Melts rich in Na₂CO₃ or sodium and magnesium carbonates can be retained



Fig. 2 Scanning electron micrograph of carbonatite melt pool in sample T-2508. This sandwich experiment was run at 22 kbar and 1,000 °C for 3 h. The carbonatite melt has quenched into large prismatic dolomite (D) and an acicular Na-Mg carbonate mineral (N). The silicate assemblage (S) consists of olivine, orthopyroxene and amphibole. Black spots in the Na-Mg quench carbonate region are due to volatiliziation during microprobe analysis.

and analysed only if sample preparation avoids all use of water because these phases are extremely soluble and mechanically soft and fragile.) The chemical composition of the carbonatite melt was determined using a layered charge or 'sandwich' technique in which 10% of an estimated melt composition was added as a basal layer to amphibole-bearing pyrolite +5% dolomite. This technique aims to produce sufficient separation of melt and residue for defocused beam or area-scan analysis of the melt, while at the same time ensuring reaction and equilibrium between melt-rich and residue-rich layers. The charge was placed in a small-capacity $\text{Ag}_{75}\text{Pd}_{25}$ capsule and surrounded by a magnesite buffer inside a large $\text{Ag}_{50}\text{Pd}_{50}$ capsule. Experiments were run at 22 kbar and 1,000 °C for 3 h. Pure Na_2CO_3 was used as the melt phase in the initial sandwich experiment. The resultant melt composition (analysed by defocused beam analysis of quench-aggregate as in Fig. 2) formed the basis of the next estimated melt. Five sandwich experiments were performed in this iterative manner to generate a melt in equilibrium with the peridotite assemblage in which the silicate mineral compositions matched precisely those observed in the (amphibole-pyrolite minus 40% olivine plus 5% dolomite) charge at the same pressure and temperature. The final peridotite assemblage consisted of olivine, orthopyroxene, clinopyroxene, pargasitic amphibole, spinel and ilmenite (Table 1). The melt produced is a carbonatite rich in Ca, Mg, Fe and Na which quenches to an assemblage of dolomite plus Na-Mg-carbonate. Our best estimate of the composition of the melt (Table 1) is derived by defocused beam analysis of the quench aggregate. Both K_2O and P_2O_5 (which were present in the (amphibole-pyrolite minus 40% olivine) layer, but were not included in the composition added as the estimated melt layer) are strongly partitioned into the melt fraction, demonstrating reaction between silicate and carbonate components.

The quenched melt is very similar texturally to samples of carbonatite described from Oldoinyo Lengai, Homa mountains, and Kaiserstuhl⁴⁻⁶. Figure 2 shows a carbonate melt pool surrounded by an ol-opx-amphibole assemblage. The melt pool contains large prismatic dolomite crystals surrounded by acicular Na-Mg carbonate grains. Low Mg^* (defined as $100 \times \text{Mg}/(\text{Mg} + \text{Fe})$) and radiating habit of the dolomite indicates that it is a quench phase and not a primary mineral. (Sub-solidus dolomite (25 kbar, 925 °C) has $\text{Mg}^* = 98$, coexisting with olivine of $\text{Mg}^* = 88$. Sub-solidus magnesite (30 kbar, 900 °C) has $\text{Mg}^* = 95$, coexisting with olivine of $\text{Mg}^* = 90$.)

The experimental carbonatite melt in equilibrium with lherzolite is more Mg-, Ca- and Fe-rich than the natro-carbonatite at Oldoinyo Lengai. Residual melts similar to the natural natro-

carbonatite could be produced from the high pressure experimental composition by olivine, spinel and dolomite fractionation (Table 1) at either mantle or crustal pressures¹⁷.

It is not necessary to invoke a process of liquid immiscibility to generate carbonatitic magmas^{18,22}. Liquid immiscibility has however been demonstrated experimentally^{14,16,19,32,33} in relevant chemical systems and could be important in generating carbonate-rich melts from primary, higher-temperature, CO_2 -rich olivine nephelinites or olivine melilitites, or from fractionated melts derived from these nephelinites (phonolites). In our experiments, it was not possible to confirm unequivocally whether there is one melt or two immiscible liquids in the above-solidus region at pressures >31 kbar, that is, beyond the stability field of amphibole. The melt fraction in the experiments quenches to a fine-grained assemblage of both silicates and carbonates, with sufficient separation of these phases to suggest immiscibility.

Our experimental study will be reported in detail, but our data set also documents the compositions of silicate minerals in equilibrium with dolomite or magnesite at sub-solidus conditions or with carbonatite melt. In particular, the Fe^{3+} solid solutions in ilmenite, spinel, amphibole, garnet and clinopyroxene in equilibrium with carbonate (EMOG buffer) provide an important calibration set for comparison with natural upper-mantle assemblages. This comparison suggests that 'normal' mantle f_{O_2} conditions are below EMOG and that peridotite- $\text{H}_2\text{O}-\text{CH}_4$ (refs 34, 35), rather than peridotite- $\text{H}_2\text{O}-\text{CO}_2$, is the relevant system for most regions of the mantle. Under local conditions of higher f_{O_2} and of heterogeneity of f_{O_2} , which could exist particularly in deep continental lithosphere or close to active or relict subducted lithosphere, the presence of carbonate and amphibole in mantle lherzolite can be expected. In these circumstances, primary carbonatite melts enriched in Na, K and P will occur under reasonable geothermal gradients and will be particularly important if these gradients locally approach the oceanic geothermal gradient. We have demonstrated an important melting reaction in complex 'fertile' peridotite compositions that considerably alters previous estimations of the peridotite- $\text{CO}_2-\text{H}_2\text{O}$ solidus, and we have shown that primary sodic dolomitic carbonatite magmas can be generated at pressures >21 kbar and $1,080 > T > 930$ °C. These carbonatite magmas may segregate at these depths and move rapidly to surface eruption, but they are extremely reactive towards their mantle wall rocks at pressures <21 kbar and may infiltrate lherzolite to produce distinctive metasomatized lherzolite or wehrlite assemblages, releasing CO_2 vapour as they do so (D.H.G. and M.E.W., submitted for publication).

We acknowledge the technical assistance of K. Harris, W. Jablonski and S. Stephens. This project was supported by a Commonwealth Postgraduate Scholarship (M.E.W.) and the Australian Research Grants Committee (D.H.G.). Constructive reviews by J. Gittins and P. J. Wyllie greatly improved the links between this work and earlier investigations on the carbonatite theme.

Received 3 May; accepted 27 July 1988.

1. Olafsson, M. & Eggler, D. H. *Earth planet. Sci. Lett.* **64**, 205-315 (1983).
2. Wyllie, P. J. *Earth planet. Sci. Lett.* **82**, 391-397 (1987).
3. Eggler, D. H. *Earth planet. Sci. Lett.* **82**, 398-400 (1987).
4. Dawson, J. B., Garson, M. S. & Roberts, B. *Geology* **15**, 765-768 (1987).
5. Deans, T. & Roberts, B. *J. geol. Soc. Lond.* **141**, 563-580 (1984).
6. Keller, J. *J. Volcan. geotherm. Res.* **9**, 423-431 (1981).
7. LeBas, M. J. *Miner. Mag.* **44**, 133-140 (1981).
8. McKie, D. & Frankis, E. J. *Z. Kristallogr.* **145**, 73-95 (1977).
9. Hay, R. L. *Geology* **11**, 599-602 (1983).
10. Deines, P. & Gold, D. P. *Geochim. cosmochim. Acta* **37**, 1709-1733 (1973).
11. Sheppard, S. M. F. & Dawson, J. B. *Phys. Chem. Earth* **9**, 747-763 (1975).
12. Nelson, D. R., Chivas, A. R., Chappell, B. W. & McCulloch, M. T. *Geochim. cosmochim. Acta* **52**, 1-17 (1988).
13. Lancelot, J. R. & Allegre, C. J. *Earth planet. Sci. Lett.* **22**, 233-238 (1974).
14. Koster Van Groos, A. F. *Am. J. Sci.* **275**, 163-195 (1975).
15. Donaldson, C. H. & Dawson, J. B. *Contr. Miner. Petrol.* **67**, 139-149 (1978).
16. Freestone, I. C. & Hamilton, D. L. *Contr. Miner. Petrol.* **73**, 105-117 (1980).
17. Twyman, J. D. & Gittins, J. in *Alkaline Igneous Rocks* (eds Fitton, J. G. & Upton, B. G. J.) 85-94 (Blackwell, Oxford, 1987).

18. Le Bas, M. J. in *Alkaline Igneous Rocks* (eds Fitton, J. G. & Upton, B. G. J.) 53-83 (Blackwell, Oxford, 1987).
19. Koster van Groos, A. T. & Wyllie, P. J. *Am. J. Sci.* **26**, 932-967 (1968).
20. Wyllie, P. J. & Tuttle, O. F. *J. Petrology* **1**, 1-46 (1960).
21. Wyllie, P. J. & Huang, W. L. *Contr. Miner. Petrol.* **54**, 79-107 (1976).
22. Wyllie, P. J. in *Magmatic Processes: Physicochemical Principles* (ed. Mysen, B. O.) 107-119 (Geochem. Soc. spec. Publ. 1, 1987).
23. Eggler, D. H. *Australian Journal of Earth Sciences spec. Publ.* **14**, IV International Kimberlite Conference (1988).
24. Eggler, D. H. *Am. J. Sci.* **278**, 305-343 (1978).
25. Brey, G. & Green, D. H. *Contr. Miner. Petrol.* **55**, 217-230 (1976).
26. Brey, G., Brice, W. R., Ellis, D. J., Green, D. H., Harris, K. L. & Ryabchikov, I. D. *Earth planet. Sci. Lett.* **62**, 63-74 (1983).
27. Green, D. H. *Earth planet. Sci. Lett.* **19**, 37-55 (1973).
28. Eggler, D. H. & Baker, D. R. in *Advances in Earth and Planetary Sciences: High Pressure Research in Geophysics 12* (eds Akimoto, S. & Manghnani, M. H.) 237-250 (Reidel, Dordrecht, 1982).
29. Holloway, J. R. *Geochim. cosmochim. Acta* **37**, 651-666 (1973).
30. Green, D. H. *Earth planet. Sci. Lett.* **17**, 456-465 (1973).
31. Green, D. H. et al. *Proc. Pacif. Rim Congr. A.I.M.M. Victoria* 621-632 (1987).
32. Hamilton, D. L., Freestone, I. C., Dawson, J. B. & Donaldson, C. H. *Nature* **279**, 52-54 (1979).
33. Wendlandt, R. F. & Harrison, W. J. *Contr. Miner. Petrol.* **69**, 409-419 (1979).
34. Taylor, W. R. & Green, D. H. *Nature* **332**, 349-352 (1988).
35. Green, D. H., Falloon, T. J. & Taylor, W. R. in *Magmatic Processes: Physicochemical Principles* (ed. Mysen, B. O.) 139-154 (The Geochemical Society, 1987).

Winter storm effects on the spawning and larval drift of a pelagic fish

David M. Checkley Jr, Sethu Raman, Gary L. Maillet & Katherine M. Mason

Department of Marine, Earth, and Atmospheric Sciences, Box 8208, North Carolina State University, Raleigh, North Carolina 27695-8208, USA

Recruitment for many marine organisms depends on survival and transport of eggs and larvae from spawning grounds to nursery areas¹. We investigated the effects of winter storms and the Gulf Stream on the spawning, development and drift of the Atlantic menhaden, *Brevoortia tyrannus*, which spawns offshore² and metamorphoses in estuaries³. Spawning was maximal during storms in water upwelled near the western edge of the Gulf Stream. Eggs and larvae drifted shoreward with abundant food in the warm surface stratum of a density-driven circulation maintained by the large sea-air heat flux. We suggest that the Atlantic menhaden and other species have evolved to reproduce in winter near warm boundary currents, including the Gulf Stream and Kuroshio, as a result of physical conditions that permit the rapid development and shoreward drift of their eggs and larvae, with consequent high recruitment and fitness.

The Atlantic menhaden spawns most intensively during winter, south of Cape Hatteras and west of the Gulf Stream² (Fig. 1). Within six days of fertilization, larvae feed on the microzooplankton⁴ and enter estuaries and metamorphose after 30-90 days³. A distinct maximum in the frequency of winter storms on the US east coast exists near Cape Hatteras; cyclones and cold fronts occur every 2-14 days^{5,6}. Survival from egg to recruitment at six months is highly variable for the Atlantic menhaden and was correlated to monthly estimates of Ekman transport at a station 56 km SE of Cape Hatteras for 1955 to 1970 (ref. 7), but not thereafter (W. Schaff, personal communication). We postulated that events on a scale of days rather than months, and processes other than Ekman transport, may be important to the spawning and larval development and drift of the Atlantic menhaden^{8,9}.

Our study was carried out during the Genesis of Atlantic Lows Experiment (GALE), a study of US east coast storms between 15 January and 15 March 1986 (ref. 10). Meteorological observations (air temperature and pressure, relative humidity, wind speed and direction, and sea-surface temperature) were made continuously at an instrumented buoy (34°10'N, 77°15'W) and from the RV *Cape Hatteras*¹¹. Oceanographic observations were made from the RV *Cape Hatteras* during 12 sampling grids

with stations of sea surface temperature (SST) between 14 and 21 °C. Two to seven major and up to 64 minor stations comprised a sampling grid. Major stations included a continuous vertical profile of temperature, salinity and density; a vertical profile of nitrate, plant pigments and zooplankton ($\geq 20\text{-}\mu\text{-mesh}$) pumped from 10-12 discrete depths; two trawls (150- $\mu\text{-mesh}$) from each of two depth strata (0-10, 10-30 m or bottom) for fish larvae; and one vertically integrated plankton tow (150- $\mu\text{-mesh}$) for fish eggs. Only the latter collection was made at minor stations. All samples were from within the upper 30 m of the water column.

The dominant meteorological sequence during GALE (Fig. 2) consisted of strong NE winds (24-26 January), passage of a low-pressure system (26-27 January), and a strong cold air outbreak (CAO, 27-30 January). This sequence was followed by a 28-day period of relative calm (31 January to 27 February), punctuated by a brief period of northerly winds (13-14 February). Heat flux at the buoy was maximal ($>900\text{ W m}^{-2}$) during the CAO and averaged $145 \pm 14 (\bar{x} \pm \text{s.e.})\text{ W m}^{-2}$ during GALE; latent heat averaged 2.2 times sensible heat flux. Wind speed and stress were greatest during the 27-30 January CAO (13 m s^{-1} , 4.9 dyn cm^{-2} , 298°). Vector averages of winds at the buoy during GALE were, for speed, 1.4 m s^{-1} from the WNW (301°) and, for stress, 0.42 dyn cm^{-2} from the NW (317°). Spectral analysis indicated dominant periods of 7.2 days for barometric pressure and 4.1 days for heat flux and wind stress during GALE, consistent with a cycle of NE winds, cyclone passage, and NW winds (Fig. 1).

Northeast winds were coincident with or followed by increased concentrations of dissolved nitrate offshore (Fig. 2). Nitrate exceeded $2\text{ }\mu\text{M}$ near the surface ($\leq 10\text{ m}$) only offshore in '18° water', a water mass characteristic of the cool western wall of the Gulf Stream¹². Chlorophyll concentration ($\mu\text{g l}^{-1}$) increased from 1.04 ± 0.03 ($n=106$) before the late January storm to 1.28 ± 0.07 ($n=91$) 1-3 days after the late January storm as a result of resuspension of settled phytoplankton, and to 1.75 ± 0.18 ($n=35$) 11-12 days post-storm as a result of growth as nitrate was used by the phytoplankton. Similar events followed NE winds in early March. Our observations are consistent with the prediction that NE winds cause upwelling at the western edge of the Gulf Stream¹³.

The cross-shelf circulation between the Gulf Stream and Mid-Shelf Fronts (GSF and MSF, respectively) during GALE was shoreward near the surface and offshore at depth (Fig. 1). During each sampling grid, warm water overlaid cool water and both strata were of oceanic salinity ($\geq 35.5\text{ g kg}^{-1}$). The mean pycnocline depth was 15.1 ± 1.0 ($n=52$). Physical stability, the density difference between depth ($\leq 30\text{ m}$) and surface divided by the depth, was high on average ($1,108 \pm 131 \times 10^{-8}\text{ m}^{-1}$, $n=52$), maximal for stations with $17.5\text{-}18.5^\circ\text{C}$ SST ($2,074 \pm 228 \times 10^{-8}\text{ m}^{-1}$, $n=11$), and unaffected by storm passage. Surface salinity increased as temperature decreased from the GSF to the MSF ($-0.012 \pm 0.002\text{ g kg}^{-1}\text{ }^\circ\text{C}^{-1}$, $n=450$), indicating the occurrence of evaporative cooling. Assuming an average heat flux of 145 W m^{-2} and a 15 m thickness for the surface stratum, we calculate its mean shoreward flow to have been 2.4 cm s^{-1} . Thus, we envision a circulation in which nutrient-rich 18° water upwells at the GSF, moves shoreward near the surface with loss of heat and nutrients, sinks at the MSF, and flows offshore at depth (Fig. 1). This circulation, though episodic in intensity, persisted throughout GALE, primarily because of the interaction of cool, dry northerly winds with shelf waters warmed by the Gulf Stream, and despite the opposing average wind stress from the NW.

The distribution of the zooplankton was consistent with this circulation. Vertical aggregation of the microzooplankton available to first-feeding larvae of the Atlantic menhaden was greatest in physically stable water (that is, shoreward of upwelling at the GSF): the variance/mean ratio for the summed concentrations of rotifers, tintinnids and copepod nauplii with 40-160 μm

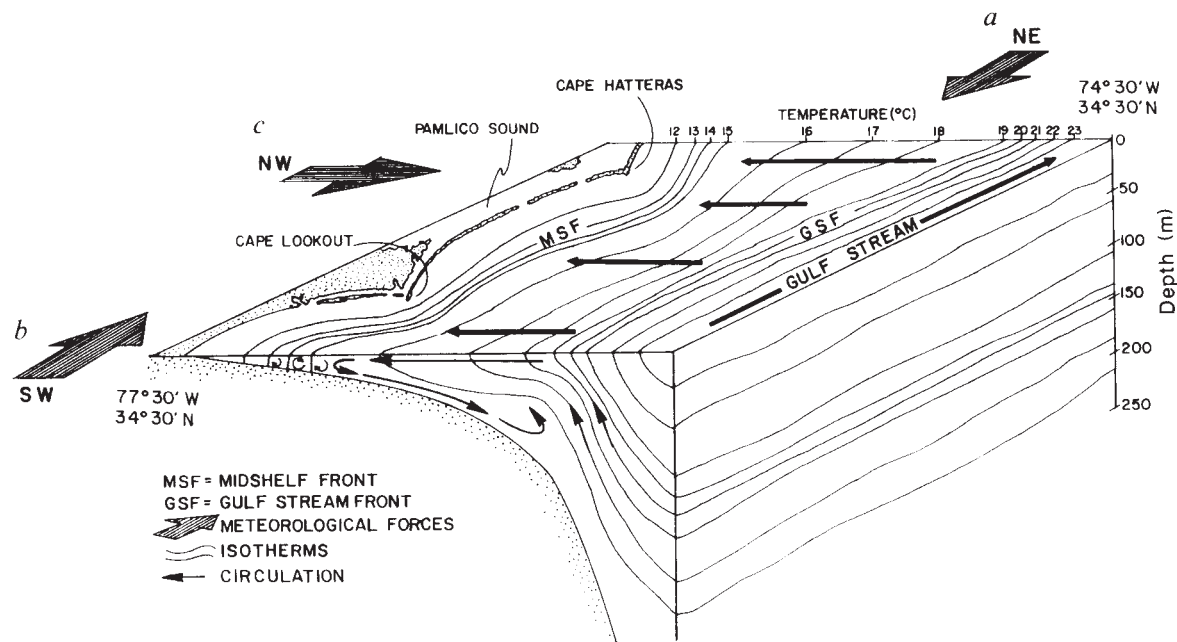


Fig. 1 Wintertime meteorological forces and cross-shelf circulation off northern North Carolina, region of greatest spawning of the Atlantic menhaden². Winter weather cycle consists of *a*, cool moist NE winds; *b*, cyclone passage with warm moist SW winds; and *c*, cold air outbreak with cold dry NW winds. NE winds cause upwelling at GSF¹³. Cross-shelf circulation is maintained by sinking near MSF and oceanward bottom flow of water cooled by heat loss to atmosphere, and replacement by shoreward surface flow from GSF.

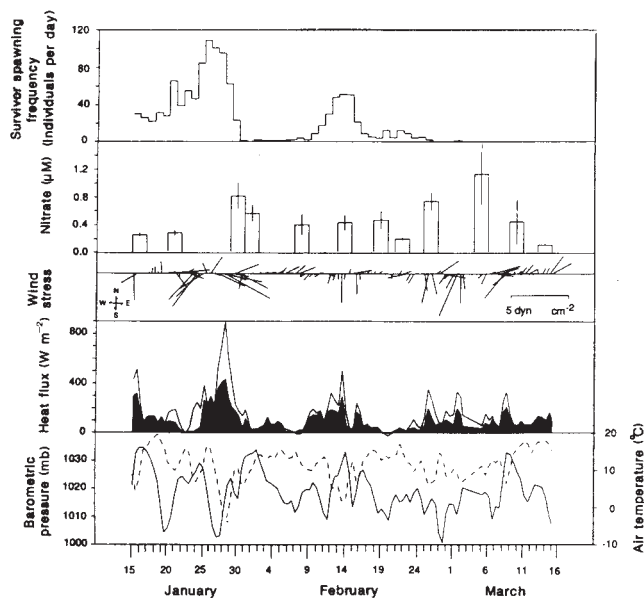


Fig. 2 Meteorological and oceanographic conditions during GALE. Barometric pressure (—), air temperature (---), heat flux (shaded, latent; line, sensible+latent), and wind stress (vectors point downwind) are 12-h averages of hourly values at 34°10'N, 77°15'W. Fluxes of latent and sensible heat and wind stress calculated using iteration method²⁶. Nitrate is average ($\pm$ s.e.) for upper 10 m of each sampling grid. Survivor spawning date distribution is for Atlantic menhaden larvae in GALE collections and estimated from the number of growth increments in sagittal otoliths. The first major increment in sagittae is formed at first feeding ($\sim$ 4 days after fertilization) and one increment is formed daily thereafter; hence, spawning date = collection date - (4 + number of increments). Distribution of eggs and small larvae and rapid changes in spawning date frequency indicate rapid and large variations in spawning rather than egg and larva mortality. Spawning was greatest during periods of high wind stress in January and February; spawning has not been observed in this region in March².

maximal width was positively correlated with stability (Spearman's rank correlation, $r = 0.39$, $P < 0.005$, $n = 52$). The average depth of maximal abundance of this microzooplankton was 10.7 ± 1.2 m ($n = 52$) and thus within the upper stratum. Genera of rotifers (*Synchaeta* and *Trichocerca*)^{14,15} and copepods (*Paracalanus*, *Clausocalanus* and *Temora*)¹⁶ typical of estuarine and coastal waters occurred in oceanic water after storms. Copepod nauplii increased in abundance from 27 ± 1 ($n = 107$) to 40 ± 2 ($n = 90$) nauplii per litre with passage of the late January storm, with post-storm concentrations as high as 125 nauplii per litre. Similar increases occurred with passage of the early March storm and copepod nauplii were, on average, abundant (36 ± 2 nauplii per litre) between the MSF and GSF during GALE. Thus in this region, winter storms and their associated offshore upwelling and cross-shelf circulation stimulate the midshelf production and vertical aggregation of the microzooplanktonic food of larval Atlantic menhaden on the scales of days and metres. First-feeding northern anchovy, *Engraulis mordax*, in contrast, require aggregations of slow-moving dinoflagellates which are dispersed by storms and upwelling^{17,18}.

The eggs and larvae of the Atlantic menhaden benefit from this circulation by drifting shoreward during winter in warm water with abundant food. Atlantic menhaden eggs float¹⁹ and Atlantic menhaden larvae were 1.5 times more abundant (individuals per square metre) near the surface (0–10 m) than deeper down (10–30 m or bottom). Rapid development due to warm temperature and abundant food in the surface stratum shortens the exposure of eggs and larvae to predators and thus enhances their survival. Otolith analysis (Fig. 2) and the offshore occurrence of eggs (21–22 January 1986) and small (≤ 5 -mm standard length) larvae (30–31 January 1986) indicate that spawning occurred near the GSF and was maximal during the late-January storm. The production, growth and shoreward transport of this cohort to the MSF are apparent when larval fish size is expressed in relation to SST and collection date (Fig. 3). This cohort moved shoreward between the GSF and MSF at ~ 2.0 cm s⁻¹, in close agreement with the shoreward transport speed estimated from the heat flux, thermocline depth, and cross-shelf temperature gradient. The largest menhaden larvae,

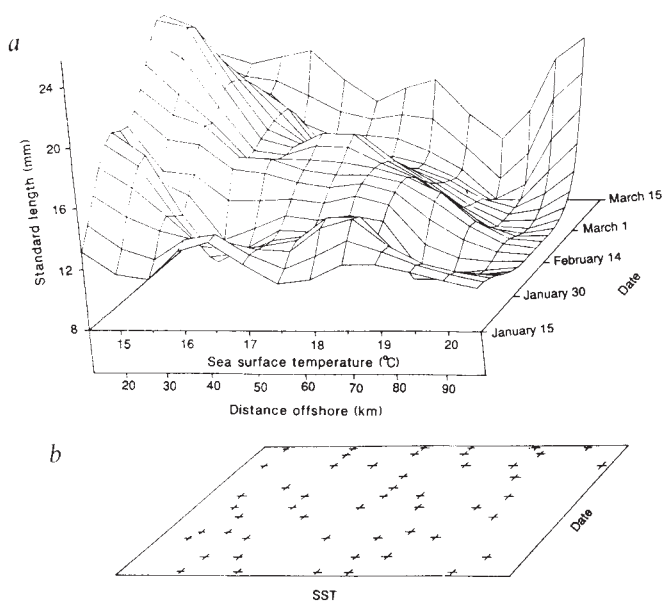


Fig. 3 Mean standard length of larval Atlantic menhaden, corrected for shrinkage, in relation to collection date (1986) and SST (mean standard length at grid nodes (a) interpolated²⁷ from measured values at 51 stations (b)). Approximate distance offshore (DO), estimated from regression of GALE data, $DO = 9.37 \text{ SST} - 108$ ($n = 241$, $r = 0.81$), shown for comparison. Maximal spawning, inferred from occurrence of eggs and analysis of otoliths (Fig. 2, top), preceded collection of smallest larvae (18°C SST, 3 February, 9.5 ± 0.3 mm average length ($n = 57$), 4.7 mm minimum length). Resultant cohort, manifest as 'ridge' of temporally increasing standard length and decreasing SST, developed and drifted shoreward in surface stratum, cooled by persistent heat loss to atmosphere, at $\sim 2 \text{ cm s}^{-1}$, consistent with rate (2.4 cm s^{-1}) estimated from measured heat flux, pycnocline depth, and cross-shelf temperature gradient. Thus, large larvae collected at 15°C SST in early March appear to have resulted from the late January spawn in 18 to 20°C SST offshore waters. Menhaden larvae of 12 – 17 mm standard length were found offshore at 19.9 and 20.4°C SST in mid-March; such larvae are presumed to die if not transported to the MSF.

apparently of this cohort, were collected on 2 May at 15°C SST, 6 km offshore. Thus, this circulation brings menhaden larvae to within ~ 20 km of shore on average and much closer at times. Subsequent movement of menhaden larvae to and through the inlets is believed to result from their vertical movement combined with the nearshore and estuarine circulations^{20,21}.

Our results indicate that the Atlantic menhaden has evolved to reproduce under physical conditions optimal for the survival and shoreward transport of its eggs and larvae. These include storms, during which upwelling and spawning occur, and persistent heat loss and stratification, during which rapid development and shoreward transport occur. Spot (*Leiostomus xanthurus*), croaker (*Micropogonias undulatus*), flounder (*Paralichthys* spp.), striped mullet (*Mugil cephalus*) and pinfish (*Lagodon rhomboides*) also spawn south of Cape Hatteras and west of the GSF in winter and have estuarine nursery areas^{3,9,22,23}. The Japanese sardine (*Sardinops melanosticta*), which supports the world's second largest fishery²⁴, spawns in winter at the western edge of the Kuroshio and its larvae occur in dense aggregations nearshore, supporting the shirasu ('white children') fishery²⁵. We suggest that each of these species has evolved to spawn during winter shoreward of a warm boundary current as a result of physical conditions, especially storm-induced upwelling and buoyancy-driven transport, which make possible the rapid development and drift of their eggs and larvae and hence enhanced recruitment and fitness. Variation in these physical

conditions may explain significant interannual variation in recruitment of these fish.

We dedicate this paper to Reuben Lasker, who died on 12 March 1988. We thank the captain and crew of the RV *Cape Hatteras* for enabling us to work at sea; C. A. Miller, A. F. Akkarapuram, R. Qiñones and A. J. Riordan for assistance in sampling and data analysis; R. S. Stemberger for identification of rotifers; the Beaufort Laboratory of the US National Marine Fisheries Service for rearing facilities and for Atlantic menhaden eggs for estimating shrinkage of larvae, and D. H. Cushing, L.-Y. Oey and A. L. Shanks for useful comments on the manuscript. Supported by the NSF.

Received 22 April; accepted 9 August 1988.

1. Cushing, D. H. *Marine Ecology and Fisheries* (Cambridge University Press, 1975).
2. Judy, M. H. & Lewis, R. M. *Distribution of Eggs and Larvae of Atlantic Menhaden, Brevoortia tyrannus, along the Atlantic Coast of the United States* (NMFS SSRF-774, Technical Report NOAA, Washington DC, 1983).
3. Hildebrand, S. F. & Cable, L. E. *Bull. U.S. Bur. Fish.* **46**, 382–488 (1930).
4. Powell, A. B. & Phonlor, G. *Fish. Bull. U.S.* **84**, 991–995 (1986).
5. Colucci, S. J. *Bull. Am. Met. Soc.* **57**, 548–553 (1976).
6. Weisburg, R. H. & Pietrafesa, L. J. *J. geophys. Res.* **88**, 4593–4610 (1983).
7. Nelson, W. R., Ingham, M. C. & Schaff, W. E. *Fish. Bull. U.S.* **75**, 23–41 (1977).
8. Yoder, J. A. *Estuar. coast. Shelf Sci.* **17**, 637–650 (1983).
9. Miller, J. M., Reed, J. P. & Pietrafesa, L. J. in *Mechanisms of Migration in Fishes* (eds McCleave, J. D., Arnold, G. P., Dodson, J. J. & Neill, W. H.) 209–225 (Plenum, New York, 1984).
10. Dirks, R. A., Kuettner, J. P. & Moore, J. A. *Bull. Am. met. Soc.* **69**, 148–160 (1988).
11. Raman, S. & Riordan, A. J. *ibid.* **69**, 161–172 (1988).
12. Pickard, G. L. & Emery, W. J. *Descriptive Physical Oceanography* (Pergamon, New York, 1982).
13. Oey, L.-Y. *J. phys. Oceanogr.* **16**, 1121–1131 (1986).
14. Ridder, M. de *Hydrobiologia* **85**, 209–225 (1981).
15. Egloff, D. A. *Hydrobiologia* **157**, 129–142 (1988).
16. Raymont, J. *Plankton and Productivity in the Oceans* Vol. 2 (Pergamon, Oxford, 1983).
17. Lasker, R. *Rapp. P.-v. Reun. Cons. perm. int. Explor. Mer* **178**, 375–388 (1981).
18. Peterman, R. M. & Bradford, M. J. *Science* **235**, 354–356 (1987).
19. Ferraro, S. P. *Fish. Bull. U.S.* **77**, 943–949 (1980).
20. Lewis, R. M. & Wilkens, E. P. H. *Chesapeake Sci.* **12**, 185–187 (1971).
21. Fore, P. L. & Baxter, K. N. *Trans. Am. Fish. Soc.* **101**, 729–732 (1972).
22. Weinstein, M. P. *Science* **214**, 351 (1981).
23. Warlen, S. M. & Chester, A. J. *Fish. Bull. U.S.* **83**, 587–599 (1985).
24. Anonymous *Fisheries Statistics: 1985* (Food and Agricultural Organization of United Nations, Rome, 1987).
25. Funakoshi, S. & Yanagibush, S. *Bull. Jap. Soc. Fish. Oceanogr.* **44**, 29–42 (1983).
26. Liu, W. T., Katsaros, K. B. & Businger, J. A. *J. atmos. Sci.* **36**, 1722–1735 (1979).
27. Akima, H. *ACM Trans. math. Software* **4**, 148–159 (1978).

Direct use of high molecular weight polysaccharide by heterotrophic flagellates

Evelyn B. Sherr

University of Georgia Marine Institute, Sapelo Island, Georgia 31327, USA

In aquatic ecology it is assumed that the heterotrophic (non-photosynthetic) protozoa that are abundant in marine and fresh water cannot compete with smaller-sized bacteria for labile dissolved substrates of low relative molecular mass (M_r), such as simple sugars and amino acids, at the low concentrations found *in situ*^{1,2} (typically 1 – $100 \mu\text{g litre}^{-1}$) because of surface to volume considerations. Consequently dissolved organic matter (DOM) is not considered to be a source of nutrition for heterotrophic protozoa, which are thought to be predominantly phagotrophic on bacteria and other small planktonic cells². But there is no information as to whether protozoa might be able to use DOM of higher M_r . Here I report that heterotrophic flagellates in a salt marsh estuary and in a freshwater pond were able to ingest molecules of the polysaccharide dextran of $M_r > 500,000$ (500K). Mixed species cultures of estuarine and pond flagellates also showed enhanced growth in the presence of 2,000K, but not 40K, dextran. Compounds with high- M_r are often a large fraction of total DOM in natural waters^{3,4}, and the concentrations (dry weight litre^{-1}) of some types of labile high- M_r compounds in natural waters are equal to, or

greater than, those of bacterioplankton^{5,6}. Thus DOM of high M_r may be an alternative food resource for aquatic flagellates. Direct ingestion of high M_r compounds by heterotrophic flagellates would represent a more efficient pathway for returning a portion of DOM to aquatic food webs compared with the DOM-bacteria-flagellate-ciliate microbial loop⁷.

Samples containing natural assemblages of protozoa were collected from the surface waters of the salt marsh estuary adjacent to Sapelo Island, Georgia, and from a freshwater pond on the south end of the island. For dextran-uptake experiments, samples were prescreened through 15- μ m-mesh Nitex netting to remove larger predators (such as microplanktonic ciliates and metazoans). Experiments were run either within one hour of sampling or the next day, to allow flagellate populations to increase after removal of predators. Results of five experiments, three with estuarine water and two with pond water, are presented in Table 1. From 6 to 20% of flagellates in individual assemblages were found to have ingested dextran-FITC (fluorescein isothiocyanate) of 2,000K and 2–17% to have ingested dextran-FITC of 500K. There was no evidence for flagellate ingestion of dextran-FITC of $M_r \leq 150$ K in any of the samples.

Flagellates observed with ingested dextran were typically 2 to 10 μ m monads and choanoflagellates⁸ (Fig. 1). As shown in the example in Fig. 1, the ingested dextran molecules appeared as amorphous fluorescence filling individual food vacuoles in the cytoplasm of flagellates. Although most of the flagellates with dextran-FITC also appeared to be bacterivorous, judging from the presence of bacterial-sized diamidinophenylindole (DAPI)-stained inclusions in vacuoles, the dextran fluorescence itself did not appear to be associated with ingested bacteria.

Table 1 Per cent of flagellates containing dextran-FITC

Sample	Heterotrophic flagellates number per ml	% Flagellates with dextran-FITC of M_r fractions		
		≤ 150 K	500 K	2,000 K
Estuary, high tide	1,870	0	4	12
Estuary, low tide	2,330	0	2.2	6.3
Estuary, low tide, water held overnight	3,810	0	11	14
Pond	1,243	0	17	20
Pond, water held overnight	1,364	0	3	11

Abundances of heterotrophic flagellates and percentages of heterotrophic flagellates containing dextran-FITC fluorescence after two-hour incubations of estuarine and pond water amended with 1 mg litre⁻¹ of dextran molecules of different M_r s. 100 ml subsamples of the screened water were added to 400 ml Whirl-pak bags which had been leached overnight with 10% HCl and then copiously rinsed with deionized water. Dextran, a linear polysaccharide produced by the bacterium *Leuconostoc mesenteroides*, was labelled with fluorescein isothiocyanate (FITC) (Sigma), and samples with M_r 20K, 40K, 70K, 150K, 500K and 2,000K were added to separate subsamples of water to give a final concentration of 1 mg dextran-FITC per litre, equivalent to 0.4 mg carbon per litre, for each of the fractions of different M_r . Freshly collected samples were incubated for 2 h at 15°C, which was near ambient *in situ* temperature, in the dark; samples held overnight were incubated at room temperature for 2 h. After incubation, a 10 ml aliquot of each subsample was fixed with a Lugol/formalin decolouration technique, which prevents preservation-induced egestion of the contents of protozoan food vacuoles (F. Rassoulzadegan *et al.*, manuscript in preparation). Protozoa in the samples were stained with diamidinophenylindole (DAPI) (ref. 19) and filtered on to 0.8 μ m Nuclepore-black filters. After mounting on to slides, the filters were examined at 1,250 $\times$ magnification using a Zeiss Universal Epifluorescence Microscope with a 75 W xenon lamp and Zeiss filter sets 47 77 02 (for DAPI fluorescence) and 47 77 09 (for FITC fluorescence). Heterotrophic flagellates were located and enumerated using the DAPI filter set, and each flagellate observed was inspected for the presence of dextran-FITC in its food vacuoles by switching filter sets. 100–200 fields (from 50 to 100 heterotrophic flagellate cells) were examined for each sample.

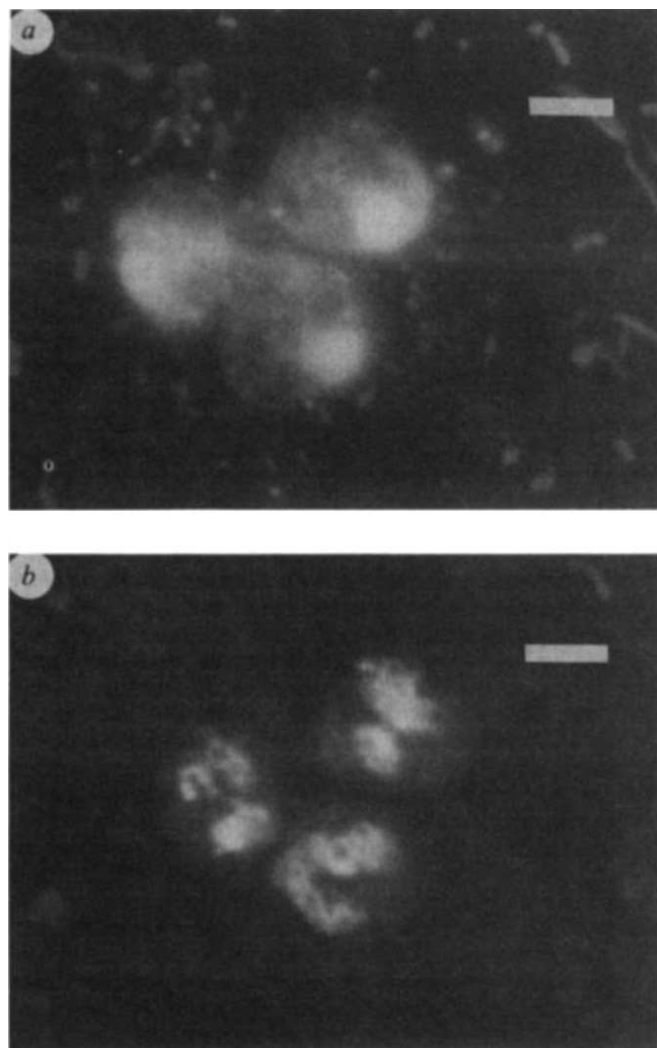


Fig. 1 Two individual cells of a freshwater, stalked choanoflagellate, *Codosiga* sp. *a*, Visualized by DAPI-fluorescence, showing the cytoplasm and flagella of the cells. The bright region of each cell is the nucleus, which stains most intensely with DAPI. *b*, The same cells seen in FITC-fluorescence, showing amorphous accumulation of 2,000K dextran-FITC molecules in the food vacuoles of the cells. Scale bars 5 μ m.

A few bacteria, mostly larger forms, exhibited FITC-fluorescence over the entire range of dextran additions, but the flagellates had FITC-fluorescence in their vacuoles only in the two additions of dextran of highest M_r . These observations indicated that the fluorescence observed in the flagellates resulted from direct ingestion of dextran molecules, and not from uptake of bacterial-associated dextran-FITC.

To test whether flagellates could use the ingested dextran for growth, I ran three experiments in which the increase in numbers of heterotrophic flagellates during 18 h incubations was compared in treatments with or without added dextran. In the first experiment, growth of estuarine flagellates on 2,000K dextran was compared in treatments with and without the addition of prokaryotic inhibitors to prevent bacterial growth²⁰. In the other two experiments, growth of estuarine and pond flagellates was compared in treatments with either 40K dextran or 2,000K dextran, with prokaryotic inhibitors added to all treatments. Flagellate numbers were significantly higher ($\alpha = 0.1$) compared with the control in all treatments with 2,000K dextran, whereas flagellates showed no enhanced growth with addition of 40K dextran (Table 2). Also, in the first experiment there was no

Table 2 Heterotrophic flagellate growth response to dextran

Treatment	Time (h)	Flagellates (10^3 cells ml^{-1})		
		Estuarine culture I	Estuarine culture II	Pond culture
All	0	8.9 $\pm$ 0.2 ni	7.2 $\pm$ 0.5	7.7 $\pm$ 0.2
No addition	18	11.6 $\pm$ 1.1 ni	5.9 $\pm$ 0.1	11.9 $\pm$ 0.9
40K dextran	18	—	5.1 $\pm$ 0.2	9.5 $\pm$ 1.0
2,000K dextran	18	19.1 $\pm$ 2.0*	9.3 $\pm$ 0.2*	22.0 $\pm$ 0.8*
2,000K dextran	18	18.2 $\pm$ 1.3* ni		

ni, No prokaryotic inhibitor added. Comparison of the increase in numbers of heterotrophic flagellates in separate treatments with either no added polysaccharide or with 4 mgC litre⁻¹ of 40 K dextran or 2,000K dextran. Mixed species assemblages of heterotrophic flagellates were grown up by incubation of 2 l aliquots of water filtered through 3.0 μm filters, at 26 °C for 48 h, which yielded a mixed-species culture of bacterivorous flagellates ranging from 2 to 8 μm . Before each experiment, the flagellate cultures were diluted 1:2 with culture water filtered through 0.2 μm filters) to lower the bacterial concentration, and prokaryotic inhibitors were added (200 mg litre⁻¹ vancomycin and 1 mg litre⁻¹ penicillin²⁰), to prevent bacterial growth, except as noted. All treatments were in 50 ml water incubated in acid-soaked and rinsed Whirl-pak bags at 26 °C in the dark. Flagellate counts were made as described in Table 1. Mean value of triplicate bags $\pm$ 1 standard deviation. Significant differences between treatment means were determined using the SPSS/PC ONEWAY ANOVA program and the Duncan multiple range test for multiple comparisons between means.

* Mean significantly different from other treatment means at 18 h at the $\alpha = 0.01$ level.

difference in flagellate yield between treatments with and without prokaryotic inhibitor. These two results indicated that the enhanced flagellate growth followed direct ingestion of the polysaccharide, rather than being a consequence of ingestion of bacteria growing on the added dextran. Of course the flagellates must have also ingested other materials with the dextran, probably bacterial cells, to obtain mineral nutrients and other substances necessary for their reproduction, such as specific amino acids or vitamins.

I also ran two analogous experiments comparing bacterial growth on glucose, on 40K dextran, and on 2,000K dextran. Estuarine bacteria showed enhanced growth in response only to glucose, and although pond bacteria did show a significant response ($\alpha = 0.05$) to dextran, there was no difference in pond bacterial growth on 40K and 2,000K dextran (Table 3).

It is known that heterotrophic protozoa are able to take up DOM because some species of both flagellates and ciliates have been grown axenically on DOM media of low M_r (refs 9 and 10). But this has been accomplished only at DOM concentrations of g litre⁻¹, representing a 10^3 -fold increase in concentration over that of the dextran added here. Experiments designed to show the uptake of DOM of low M_r by protozoa at *in situ* concentrations of substrate have produced negative results^{1,11}. No previous study, however, has tested the ability of free-living protozoa to incorporate compounds of high M_r when protozoa could have an advantage over bacteria. Bacteria must degrade such compounds extracellularly to monomer or dimer fragments before they can be taken into the cell, but protozoa could ingest entire polymeric molecules for subsequent internal digestion. The results of the flagellate and bacteria growth experiments (Tables 2 and 3) support this idea: flagellates showed enhanced growth on 2,000K dextran, estuarine bacteria showed no increased growth, and pond bacteria showed only moderate growth on dextran when compared with bacterial growth on the same concentration of glucose.

Although the dextran molecules ingested by the flagellates were in the size range of the larger-sized fractions of DOM found in natural water^{3,4}, they are orders of magnitude smaller than the smallest particulate food items (bacteria) ingested by free-living protozoa. An average aquatic bacterium has a dry

Table 3 Bacterial growth response to glucose and dextran

Treatment	Time (h)	Bacterial (10^6 cells ml^{-1})	
		Estuarine water	Pond water
All	0	4.2 $\pm$ 0.1	4.9 $\pm$ 0.2
No addition	18	7.9 $\pm$ 0.5	17.9 $\pm$ 0.9
Glucose	18	14.0 $\pm$ 0.5*	36.9 $\pm$ 2.5*
40K dextran	18	7.6 $\pm$ 0.1	25.6 $\pm$ 0.3†
2,000K dextran	18	8.0 $\pm$ 0.5	25.9 $\pm$ 1.9†

Comparison of the increase in bacterial numbers in 0.8 μm -filtered estuarine and pond water in separate treatments with either no addition of organic substrate, or of amendments with 4 mgC litre⁻¹ of glucose, 40K dextran, or 2,000K dextran. Each treatment was in 100 ml aliquots of water incubated in acid-soaked and rinsed Whirl-pak bags at 26 °C in the dark. Addition of 20 μM NH_4^+ litre⁻¹ and 2 μM PO_4^{3-} litre⁻¹ to each treatment ensured that bacterial growth would not be limited by availability of nitrogen and phosphorus. Bacteria counts were made by the acridine-orange direct-count method²¹. Mean value of triplicate samples $\pm$ 1 standard deviation. Statistical analyses as Table 2.

* Mean significantly different from other treatment means at 18 h at the $\alpha = 0.01$ level.

† Mean significantly different from other treatment means at 18 h at the $\alpha = 0.05$ level.

weight of 40 fg per cell¹², which can be translated into an M_r equivalent of 2.4×10^{10} K. Thus a 2,000K polysaccharide molecule is 10^7 times smaller than a bacterial cell by mass. This opens the possibility that heterotrophic flagellates could use organic materials spanning this entire size range, which includes colloidal particles of 1 μm to 0.001 μm (ref. 13).

High M_r organic matter (>100K) is often a significant fraction of total DOM in natural waters. Ogura⁴ has reported that compounds of high M_r range from 0.1 to 1.5 mgC litre⁻¹ in Tokyo Bay and account for 8–45% of the total DOM. Wheeler³ has reported DOM of high M_r at concentrations of 0.1 to 1.8 mgC litre⁻¹ in Georgia estuarine waters. The high- M_r fraction of DOM is known to contain readily degradable compounds. Wheeler³ found concentrations of 0.04 to 0.3 mgC litre⁻¹ of carbohydrate in the >100K fraction of estuarine DOC. Burney *et al.*⁵ measured total polysaccharide concentrations of 0.3 to 1.0 mgC litre⁻¹ in waters of a Rhode Island salt marsh. These concentrations are equal to, or greater than, the concentration of bacterial biomass in coastal waters, which generally ranges from 0.02 to 0.15 mgC litre⁻¹ (ref. 14). Other high- M_r compounds occurring in natural waters that could be ingested by flagellates include DNA⁶ and mucopolysaccharides¹⁵.

The present theory of carbon flow in marine food webs holds that both particulate and dissolved non-living organic matter is degraded by bacteria⁷. It is thought that a two- or three-step trophic pathway, the microbial loop, bacteria-flagellates-ciliates, is necessary for the return of a portion of this detrital carbon to the metazoan food web⁷. Because of respiratory losses at each trophic step, microbially processed detrital carbon is hypothesized to be a minor input into higher trophic levels of most aquatic food webs¹⁶. But as heterotrophic flagellates represent an important food resource for aquatic microcrustaceans^{17,18}, direct ingestion of high- M_r DOM by flagellates would represent a short-circuit of the microbial loop, leading to a greater efficiency of transfer of non-living organic matter to metazoan food webs.

I thank Eileen Hedick for preparing the figure and Dr Barry Sherr for help during the study and in preparation of the manuscript. This work was supported by NSF and the Sapelo Island Research Foundation.

Received 22 June; accepted 10 August 1988.

1. Haas, L. W. & Webb, K. L. *J. exp. mar. Biol. Ecol.* **39**, 125–134 (1979).
2. Fenchel, T. *The Ecology of Protozoa* (Science Tech. Madison, Wisconsin, 1987).
3. Wheeler, J. R. *Limnol. Oceanogr.* **21**, 846–852 (1976).
4. Ogura, N. *Mar. Chem.* **5**, 535–549 (1977).
5. Burney, C. M., Johnson, K. M. & Sieburth, J. McN. *Mar. Biol.* **63**, 175–187 (1981).

6. DeFlaun, M. F., Paul, J. H. & Jeffrey, W. H. *Mar. Ecol. Prog. Ser.* **38**, 65-73 (1987).
7. Azam, F. *et al. Mar. Ecol. Prog. Ser.* **10**, 257-263 (1983).
8. Lee, J. J., Hutner, S. Y. & Bovee, E. C. (eds) *An Illustrated Guide to the Protozoa* (Allen, Lawrence, Kansas, 1985).
9. Sleight, M. A. *The Biology of Protozoa* (Arnold, London, 1979).
10. Rassmussen, L. & Oria, E. *Science* **190**, 464-465 (1975).
11. Glaser, D. *Microb. Ecol.* **15**, 189-201 (1988).
12. Lee, S. & Fuhrman, J. D. *Appl. envir. Microbiol.* **53**, 1298-1303 (1987).
13. Vold, R. D. & Vold, M. J. in *Encyclopedia of Chemistry* (eds Clark, G. L. & Hawley, G. G.) 263-265 (Reinhold, New York, 1966).
14. Sherr, E. B., Sherr, B. F., Fallon, R. D. & Newell, S. Y. *Limnol. Oceanogr.* **31**, 177-183 (1986).
15. Smetacek, V. & Pollehn, F. *Ophelia* **26**, 401-428 (1986).
16. Ducklow, H. W., Purdie, D. A., Williams, P. J. leB & Davis, J. M. *Science* **232**, 865-867 (1986).
17. Roman, M. R. *et al. Mar. Ecol. Prog. Ser.* **42**, 39-52 (1988).
18. Sanders, R. W. & Porter, K. G. *EOS* **68**, 1705 (1987).
19. Porter, K. G. & Feig, Y. S. *Limnol. Oceanogr.* **25**, 943-948 (1980).
20. Sherr, B. F., Sherr, E. B., Andrew, T. L., Fallon, R. D. & Newell, S. Y. *Mar. Ecol. Prog. Ser.* **32**, 169-179 (1986).
21. Hobbie, J. E., Daley, R. J. & Jasper, S. *Appl. envir. Microbiol.* **33**, 943-948 (1977).

Evolution of a bacteria/plasmid association

Judith E. Bouma & Richard E. Lenski*

Department of Ecology and Evolutionary Biology,
University of California, Irvine, California 92717, USA

Associations between bacteria and their accessory elements (viruses, plasmids and transposons) range from antagonistic to mutualistic^{1,2}. A number of previous studies have demonstrated that plasmid carriage reduces bacterial fitness in the absence of selection for specific functions such as antibiotic resistance³⁻¹³. Many studies have demonstrated increased fitness of evolving microbial populations in laboratory environments^{3,8,14-21}, but we are aware of only one study in which fitness gains were partitioned between a plasmid and its host⁸. Here, we examine the evolution of an association between a plasmid and its bacterial host. Carriage of the non-conjugative plasmid pACYC184 initially reduced the fitness of *Escherichia coli* B in the absence of antibiotic. We then cultured plasmid-bearing bacteria for 500 generations in the presence of antibiotic. The fitness of each combination of host and plasmid, with and without the culture history, was determined by competing it against a baseline strain. The results indicate adaptation by the host genome, but no plasmid adaptation. We also competed the evolved host, transformed with the baseline plasmid, against its isogenic plasmid-free counterpart. The plasmid now increased the fitness of its host.

E. coli B^{22,23} was transformed with the 4-kilobase (kb) non-conjugative plasmid pACYC184, which encodes resistance to chloramphenicol and tetracycline^{24,25}. We cultured the plasmid-bearing bacterial population for 75 days in a glucose-limited minimal salts medium²² supplemented with 10 µg ml⁻¹ chloramphenicol, which selected against plasmid-free segregants. Each day, 0.1 ml of culture was transferred into 9.9 ml of fresh medium; the cultures were maintained in 50-ml Erlenmeyer flasks at 37 °C and 120 revs per minute. The 100-fold daily increase in cell numbers corresponds to 6.64 doublings, so that 75 days of culture represented approximately 500 generations of binary fission.

To allow us to distinguish different strains in our assays of relative fitness, we used an arabinose marker. The parental *E. coli* B strain is unable to grow on arabinose. We selected mutants capable of using arabinose, which served as reference competitors. Ara⁺ strains form white colonies on tetrazolium arabinose plates [tryptone agar supplemented with arabinose (L-isomer) and triphenyltetrazolium chloride (TTC)]. The derivations of the strains used in this study are described in Table 1.

To determine the effects of plasmid carriage on host fitness

Table 1 Bacterial strains

Strain	Relevant characteristics*
JB11	plasmid-free <i>Escherichia coli</i> B
JB12	spontaneous Ara ⁺ mutant of JB11†
JB13	transformant of JB11 using plasmid pACYC184
JB14	spontaneous Ara ⁺ mutant of JB13†
JB38	derivative of JB13 cultured for 500 generations
JB44	spontaneous plasmid-free segregant of JB38‡
JB54	transformant of JB11 using plasmid from JB13§
JB55	transformant of JB44 using plasmid from JB13
JB56	transformant of JB11 using plasmid from JB38
JB57	transformant of JB44 using plasmid from JB38§
JB85	spontaneous Ara ⁺ mutant of JB44†
RL861	transformant of JB85 (Ara ⁺) using plasmid from JB13

All strains were stored at -80 °C immediately after their isolation to preclude further genetic changes.

* Ara⁺ indicates ability to use arabinose; all strains are Ara⁻ unless otherwise indicated.

† Isolated by plating on minimal arabinose agar.

‡ Isolated by testing clones on agar with and without supplemental antibiotics.

§ JB54 and JB57 are identical to JB13 and JB38, respectively; they were used to control for any possible effects of segregation and transformation on fitness.

before the 75 days in culture, we competed the plasmid-bearing strain JB13 against its plasmid-free counterpart JB12 in the glucose-limited minimal medium, without chloramphenicol. As a control for the arabinose marker, we also competed the two plasmid-free strains, JB11 and JB12. Each replicate experiment ran for six days, with samples plated daily. Some colonies from each sample were also tested on agar supplemented with chloramphenicol to monitor for segregants appearing during the competition experiments, but they never became a large fraction of the population. Any slight increase in the frequency of segregants weakens statistical comparisons between plasmid-bearing and plasmid-free strains, so our conclusions are conservative. The arabinose marker had no significant effect on fitness, whereas there was a significant reduction in fitness of the bacterial host associated with carriage of the plasmid in the antibiotic-free environment (Table 2).

We investigated whether the 500 generations of culture history conferred increased fitness; and, if so, whether genetic changes in the bacterial chromosome, plasmid, or both were responsible for this adaptation. All combinations of plasmid and host, with and without the culture history (JB54-JB57) were competed against JB14, the baseline plasmid-bearing strain with the alternative arabinose marker, in medium supplemented with chloramphenicol. We obtained three replicate estimates of the selection coefficient for each of the strains, as described (Table 2). Analysis of variance indicates a highly significant effect associated with the bacterial chromosome, but no significant effects associated with either the plasmid or the chromosome/plasmid interaction (Table 3). Thus, the bacteria/plasmid

Table 2 Effect of plasmid carriage on host fitness before culture history

Competitors			Mean	Standard error	t
A	B	n			
JB13	JB12	26	-0.055 per day	0.018	3.084 (P < 0.01)
JB11	JB12	30	-0.003 per day	0.006	0.500 (NS)

A control for the effect of the arabinose marker was included (JB11 versus JB12). The *n* estimates of the selection coefficient were calculated by regressing the natural logarithm of the ratio of the two competitors (A/B) against time. A positive value indicates that A is more fit than B; a negative value that B is more fit than A. NS, not significant; *t* = *t*-statistic.

* To whom correspondence should be addressed.

Table 3 Contributions of bacterial chromosome and plasmid to genetic adaptation

Source	Degrees of freedom	Sums of squares	Mean square	F
Chromosome	1	3.6725	3.6725	119.625 ($P < 0.001$)
Plasmid	1	0.0069	0.0069	0.225 (NS)
Interaction	1	0.0117	0.0117	0.381 (NS)
Error	8	0.2457	0.0307	
Total	11	3.9368		

Fixed effects model analysis was used to determine the contribution to adaptation during the 500 generations of culture history. NS, not significant; $F = F$ statistic.

association underwent significant adaptation during the 500-generation experiment; this adaptation was the result of genetic change in the bacterial chromosome, but not in the plasmid.

We then determined whether adaptation by the bacterium had altered the effect of plasmid carriage on fitness. We competed JB55, the evolved bacterium transformed with the baseline plasmid, against JB85, the isogenic plasmid-free strain, in the absence of chloramphenicol. To ensure that the arabinose marker had no effect on fitness in the evolved bacterium, we included a control in which JB44 was competed against JB85. As shown in Table 4, the arabinose marker again had no significant effect on fitness, whereas there was a highly significant enhancement of host fitness associated with plasmid carriage, even in the antibiotic-free environment, after the host had evolved in the presence of the plasmid for 500 generations. Given this surprising result, we performed another control: we competed RL861 (a transformant of JB85) against the plasmid-free JB44. The highly significant enhancement of fitness was also observed in this evolved host possessing the alternative arabinose marker (Table 4).

Our results can be summarized as follows. (1) The newly introduced plasmid was antagonistic in the absence of antibiotic. (2) The host/plasmid association underwent adaptation during 500 generations in the presence of antibiotic. This adaptation was due to genetic change by the host. (3) After 500 generations, the plasmid enhances the fitness of its host even in the absence of antibiotic. Thus, an association that was formerly mutualistic in the presence of antibiotic, but antagonistic in the absence of antibiotic, has evolved into an association that is mutualistic in both environments. (4) The genetic change responsible for the new mutualistic effect arose in the host chromosome, because the evolved bacterium experiences the benefit when transformed with the baseline plasmid.

In the one previous study where fitness gains were partitioned between a plasmid and host, Helling and co-workers⁸ observed that increased fitness was due to genetic change in the host, and not in the plasmid, but they did not determine whether the cost of plasmid carriage had changed as the result of bacterial adaptation. Several other studies have demonstrated that accessory elements can enhance bacterial fitness even in the absence of selection for specific functions such as superinfection immunity or antibiotic resistance²⁶⁻³¹. However, we believe that this is the first study in which an evolutionary transition from antagonism to mutualism has been observed directly for a bacterial accessory element.

What is the nature of the genetic change in the host that causes it to benefit from plasmid infection, where previously it was harmed? The evolved bacterial host could have become tolerant of some cost imposed by the plasmid, thereby allowing a previously existing plasmid-encoded benefit to be manifest. It is interesting that a portion of pACYC184 is derived from P15A, a non-conjugative, naturally occurring plasmid that does not code for any known phenotypic properties²⁵. It is not clear how cryptic plasmids are maintained in bacterial populations. Math-

Table 4 Effect of plasmid carriage on host fitness after culture history

Competitors		n	Mean	Standard error	t
A	B				
JB55	JB85	20	+0.171 per day	0.017	10.142 ($P < 0.001$)
JB44	JB85	20	+0.009 per day	0.005	1.784 (NS)
RL861	JB44	20	+0.106 per day	0.009	11.800 ($P < 0.001$)

A control for the effect of the arabinose marker was included (JB44 versus JB85). NS, not significant.

ematical analyses^{32,33} have shown that the conditions for such plasmids to be maintained through mobilization by conjugative plasmids are extremely restrictive. Such plasmids may carry functions that benefit their hosts, but which have not been detected because they confer no obvious phenotype. We will test this hypothesis by deleting portions of the pACYC184 genome and determining whether any particular region is essential for its fitness-enhancing effect in the evolved host.

There is an important practical implication of our results. It is often assumed that the introduction of DNA for biotechnological functions will reduce the fitness of recombinant organisms, owing to costs of carriage and expression of foreign DNA. According to this 'excess baggage' hypothesis, the deliberate release of recombinant organisms into the environment is inherently safe because these costs will prevent their unintended spread^{4,34-36}. Our results caution against uncritical acceptance of this hypothesis.

We thank Pete Atsatt, Rick Condit, Loan Duong, Ralph Evans, Sue Hattingh, Keith Justice, Bruce Levin, Judy MacDougall, Toai Nguyen, Trinh Nguyen, Mayur Patel, Quang Phan, Michael Rose, Brad Shaffer and Lone Simonsen for their valuable comments, discussion, or assistance. Support for this work was provided by grants from the National Institutes of Health (GM-37105) and the Gulf Breeze Laboratory of the US Environmental Protection Agency (CR-813401) to R.E.L.

Received 4 March; accepted 18 July 1988.

- Levin, B. R. & Lenski, R. E. in *Coevolution* (eds Futuyma, D. J. & Slatkin, M.) 99-127 (Sinauer Associates, Sunderland, Mass., 1983).
- Davey, R. B. & Reaney, D. C. in *Evolutionary Biology*, Vol. 13 (eds Hecht, M. K., Steere, W. C. & Wallace, B.) 113-147 (Plenum, New York, 1980).
- Dykhuizen, D. E. & Hartl, D. L. *Microbiol. Rev.* **47**, 150-168 (1983).
- Lenski, R. E. & Nguyen, T. T. in *Planned Release of Genetically Engineered Organisms* (eds Hodgson, J. & Sugden, A. M.) 18-20 (Elsevier, Cambridge, 1988).
- Lenski, R. E. & Bouma, J. E. *J. Bact.* **169**, 5314-5316 (1987).
- Cooper, N. S., Brown, M. E. & Caulcott, C. A. *J. gen. Microbiol.* **133**, 1871-1880 (1987).
- Lee, S. W. & Edlin, G. *Gene* **39**, 173-180 (1985).
- Helling, R. B., Kinney, T. & Adams, J. *J. gen. Microbiol.* **123**, 129-141 (1981).
- Noack, D. *et al. Molec. gen. Genet.* **184**, 121-124 (1981).
- Levin, B. R. in *Antibiotic Resistance: Transposition and Other Mechanisms* (eds Mitsuhashi, S., Rosival, L. & Krcmery, V.) 197-202 (Springer, Berlin, 1980).
- Zund, P. & Lebeck, G. *Plasmid* **3**, 65-69 (1980).
- Jones, I. M., Primrose, S. B., Robinson, A. & Ellwood, C. C. *Molec. gen. Genet.* **180**, 579-584 (1980).
- Godwin, D. & Slater, J. H. *J. gen. Microbiol.* **111**, 201-210 (1979).
- Lenski, R. E. *Evolution* **42**, 433-440 (1988).
- Paquin, C. E. & Adams, J. *Nature* **306**, 368-370 (1983).
- Chao, L., Vargas, C., Spear, B. B. & Cox, E. C. *Nature* **303**, 633-635 (1983).
- Paquin, C. E. & Adams, J. *Nature* **302**, 495-500 (1983).
- Dykhuizen, D. & Hartl, D. *Evolution* **35**, 581-594 (1981).
- Graham, J. B. & Istock, C. A. *Science* **204**, 637-639 (1979).
- Luckinbill, L. S. *Science* **202**, 1201-1203 (1978).
- Atwood, K. C., Schneider, L. K. & Ryan, F. J. *Cold Spring Harb. Symp. quant. Biol.* **16**, 345-355 (1951).
- Lenski, R. E. *Evolution* **42**, 425-432 (1988).
- Lenski, R. E. & Levin, B. R. *Am. Nat.* **125**, 585-602 (1985).
- Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning* (Cold Spring Harbor Laboratory, New York, 1982).
- Chang, A. C. Y. & Cohen, S. N. *J. Bact.* **134**, 1141-1156 (1978).
- Edlin, G., Tait, R. C. & Rodriguez, R. L. *Biotechnol.* **2**, 251-254 (1984).
- Hartl, D. L., Dykhuizen, D. E., Miller, R. D., Green, L. & DeFramond, J. *Cell* **35**, 503-510 (1983).
- Biel, S. W. & Hartl, D. L. *Genetics* **103**, 581-592 (1983).
- Edlin, G., Lin, L. & Bitner, R. *J. Virol.* **21**, 560-564 (1977).
- Lin, L., Bitner, R. & Edlin, G. *J. Virol.* **21**, 554-559 (1977).
- Edlin, G., Lin, L. & Kudrna, R. *Nature* **255**, 735-737 (1975).
- Levin, B. R. in *Antibiotic Resistance Genes: Ecology, Transfer, and Expression* (eds Levy, S. B. & Novick, R. P.) 57-70 (Cold Spring Harbor Laboratory, New York, 1986).
- Levin, B. R. & Stewart, F. M. *Genetics* **94**, 425-443 (1980).
- Regal, P. J. *Recomb. DNA Tech. Bull.* **10**, 67-85 (1987).
- Colwell, R. K., Norse, E. A., Pimentel, D., Sharples, F. E. & Simberloff, D. *Science* **229**, 111-112 (1985).
- Brill, W. J. *Science* **227**, 381-384 (1985).

Migration of young neurons in adult avian brain

Arturo Alvarez-Buylla & Fernando Nottebohm

The Rockefeller University, 1230 York Avenue,
New York, New York 10021, USA

Neurons are born in the ventricular walls of the vertebrate central nervous system. From there, the young neurons migrate to their final destinations, where differentiation occurs. Neuronal migration has been described during the ontogeny of the avian and mammalian brain. Whereas in mammals most neurogenesis occurs during early development^{1,2}, in the adult avian forebrain widespread neurogenesis continues to occur³⁻⁷. How do neurons born in adulthood reach their final destination? We report here that small elongated cells⁸, born in the ventricular zone adjacent to the lateral ventricle, differentiate into mature neurons 20–40 days later, after migrating over distances of up to 5 mm. Migration rates are highest ($28 \mu\text{m h}^{-1}$) when young neurons migrate through regions which are rich in radial glia. The adult vertebrate brain offers unique opportunities for studying factors that regulate neuronal migration, pathfinding and differentiation.

We used [³H]thymidine, a marker of DNA synthesis, to document the birthplace, migration and differentiation of neurons born in adulthood (Figs 1 and 2). Of the labelled cells of adult canaries killed 1 h after [³H]thymidine injection, 68% were in the ventricular zone (VZ)^{3,9} lining the lateral wall of the lateral ventricle. The remaining labelled cells were distributed throughout the brain parenchyma and were either glia (7.5%), endothelial cells (11.6%) or cells of uncertain identification (12.9%)^{3,8}. After one day survival, the distribution of labelled cells was very similar.

The number of labelled VZ cells increased by 63% between survival days one and three, presumably reflecting division of some VZ cells labelled on day one. The number decreased thereafter; by days 20, 30 and 40 there were only 5.2%, 3.5% and 1.5% as many labelled VZ cells as on day three (Figs 2 and 3).

After one day a new type of labelled cell could be distinguished very close to the lateral ventricle. These cells have elongated nuclei with basophilic nucleoplasm and one or two darkly staining nucleoli⁸ (Fig. 1). By day six they had increased in number and moved laterally, clustering in two cohorts, one in the lobus parolfactorius and the other in the hyperstriatum ventralis and accessorium (Fig. 2). We conclude that these are migrating cells. The number of labelled migrating cells peaked at day 20 (Fig. 3), when they had reached the most lateral parts of the telencephalon (Fig. 2), a distance of 5 mm from the ventral tip of the lateral ventricle. The increase in labelled migrating cells paralleled the disappearance of labelled ventricular zone cells (Fig. 3).

During the first six days after the last [³H]thymidine injection many of the labelled migrating cells were present in regions rich in radial glia fibres^{8,10} (Fig. 2). At higher magnification ($\times 788$), many migrating cells were seen closely apposed to a radial glia fibre (Fig. 1). Even those not visibly attached to radial glial fibres were usually

found oriented parallel to them. On survival days one, three and six, there were, respectively, 50%, 29% and 32% labelled migrating cells tightly apposed to radial glia fibres. This proportion dropped to 16% and 9% on days 15 and 20; by then labelled migrating cells had scattered and many had moved past the radial glia fibre-rich region (Fig. 2).

No labelled neurons were found before survival times of 20 days, but their number increased steadily thereafter (Figs 2 and 3). The coordinated appearance and disappearance of labelled VZ cells and migrating cells and the late appearance of labelled neurons (Figs 2 and 3) strongly suggests that in each case the earlier labelled cell type is a direct predecessor of the later type. We infer that the migrating cells are young migrating neurons. These migrating cells occur only in the telencephalon⁸, and it is only in the telencephalon that neurogenesis takes place⁶. Gliogenesis occurs throughout the adult brain. The number of labelled glia in telencephalic parenchyma remained constant at all survival times tested (for example, 20.4 and 21.9 labelled glial cells per hemisection on survival days 1 and 40 respectively). The migrating cells were negative to markers for glial fibrillary acidic protein and vimentin⁸. Gliogenesis did not seem to be related to the birth of cells in the VZ nor to their subsequent migration.

The rate of migration was estimated by assuming straight migration of five labelled young migrating neurons farthest from a ventricular wall. Because the path of a migrating cell will seldom be a straight line, the calculated rates will under-estimate the actual speed of movement. The migratory rate peaked between days 3 and 6 ($28 \mu\text{m h}^{-1}$), slowing down between days 6 and 20 ($8 \mu\text{m h}^{-1}$). This drop in the rate of dispersal from the ventricle occurred as the migrating cells entered regions with

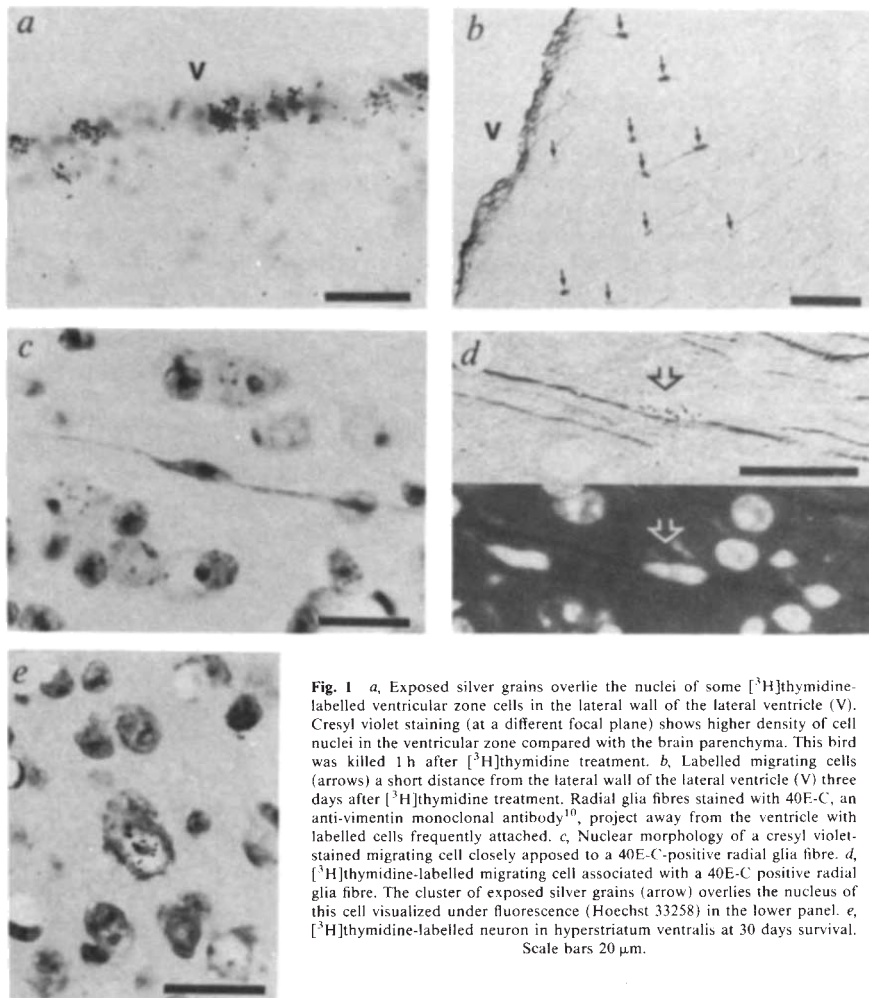
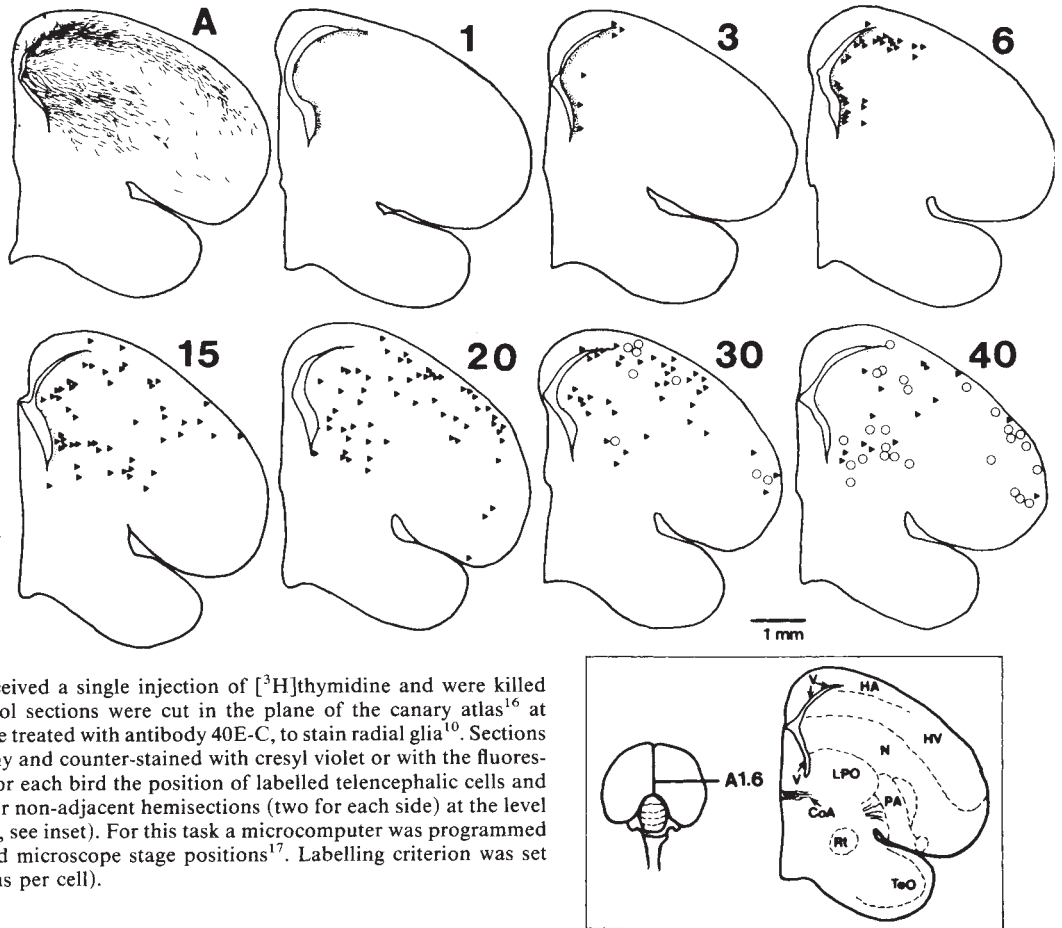


Fig. 1 *a*, Exposed silver grains overlie the nuclei of some [³H]thymidine-labelled ventricular zone cells in the lateral wall of the lateral ventricle (V). Cresyl violet staining (at a different focal plane) shows higher density of cell nuclei in the ventricular zone compared with the brain parenchyma. This bird was killed 1 h after [³H]thymidine treatment. *b*, Labelled migrating cells (arrows) a short distance from the lateral wall of the lateral ventricle (V) three days after [³H]thymidine treatment. Radial glia fibres stained with 40E-C, an anti-vimentin monoclonal antibody¹⁰, project away from the ventricle with labelled cells frequently attached. *c*, Nuclear morphology of a cresyl violet-stained migrating cell closely apposed to a 40E-C-positive radial glia fibre. *d*, [³H]thymidine-labelled migrating cell associated with a 40E-C-positive radial glia fibre. The cluster of exposed silver grains (arrow) overlies the nucleus of this cell visualized under fluorescence (Hoechst 33258) in the lower panel. *e*, [³H]thymidine-labelled neuron in hyperstriatum ventralis at 30 days survival. Scale bars 20 μm .

Fig. 2 Maps of labelled ventricular zone cells (●), migrating cells (▶) and neurons (○) in transverse brain hemisections 1, 3, 6, 15, 20, 30 and 40 days after the second of two doses of [3 H]thymidine. In map A the distribution of 40E-C-positive fibres in the telencephalon is shown⁸. The lower right inset shows a canary brain from the top and indicates the level of sectioning. Major anatomical regions at this level are CoA, anterior commissure; HA, hyperstriatum accessorium; HV, hyperstriatum ventralis; LPO, lobus parolfactorius; PA, paleostriatum augmentatum; Rt, nucleus rotundus; TeO, optic tectum; V, lateral ventricle. For this experiment and those shown in Fig. 3, twenty-one 13–15-month-old male canaries received two intramuscular injections of 2.5 μ Ci of [3 H]thymidine per gram body mass at 12-h intervals (9 a.m. and 9 p.m.). The birds were killed under deep anaesthesia in groups of three after survivals indicated above. Three additional males received a single injection of [3 H]thymidine and were killed 1 h later. Serial polyethylene glycol sections were cut in the plane of the canary atlas¹⁶ at 10 μ m intervals. Some sections were treated with antibody 40E-C, to stain radial glia¹⁰. Sections were incubated for autoradiography and counter-stained with cresyl violet or with the fluorescent DNA stain Hoechst 33258. For each bird the position of labelled telencephalic cells and their identity^{3,8} was mapped in four non-adjacent hemisections (two for each side) at the level of the CoA (A.1.6 of canary atlas¹⁶, see inset). For this task a microcomputer was programmed to store relative camera-lucida and microscope stage positions¹⁷. Labelling criterion was set at 20 $\times$ background level (5–7 grains per cell).



relatively few radial glia (Fig. 2). These conservative estimates of the migration speed are considerably faster than the 2–5 μ m h⁻¹ typical for migrating neurons in developing vertebrate brain¹¹. However, granular neurons migrating on Bergmann glia processes *in vitro* move at a speed of 33 μ m h⁻¹ (ref. 12). Our observations indicate that radial glia are important in guiding^{13,14} and hastening migration during the first week after the young neurons leave the VZ; thereafter other cues may take over¹⁵.

The number of labelled VZ cells and the number of migrating cells both exceeded the number of differentiated labelled neurons. Assuming that there had been no net gain or loss of migrating cells or neurons as a result of cells moving in or out of the plane of section, only one third of the labelled migrating cells became labelled neurons that were still present at day 40

(Fig. 3). Many labelled pycnotic cells (not shown) occurred in areas rich in migrating neurons. It seems that a majority of the migrating cells are culled before or soon after they differentiate into adult neurons.

Neuronal migration in adult avian brain seems unique in several respects. (1) The distances over which cells migrate are relatively long compared with the distances observed in developing brain. (2) The size of the brain remains fixed throughout migration. (3) The anatomical landmarks probably remain unchanged. (4) The speed at which migrating cells move away from the ventricles seems considerably faster than during development. (5) Migration through regions rich or poor in radial glia fibres permits comparisons of factors regulating pathfinding under both these conditions.

This work was supported by grants from the Whitehall Foundation and the US Public Health Service. We thank Kirk Manogue, Marta Nottebohm and Ellen Prediger for helpful editorial comments.

Received 14 April; accepted 19 August 1988.

1. Rakic, P. *Science* **227**, 1054–1056 (1985).
2. Bayer, S. A., Yackel, J. W. & Puri, P. S. *Science* **216**, 890–892 (1982).
3. Goldman, S. A. & Nottebohm, F. *Proc. natn. Acad. Sci. U.S.A.* **80**, 2390–2394 (1983).
4. Paton, J. A. & Nottebohm, F. *Science* **225**, 1046–1048 (1984).
5. Burd, G. D. & Nottebohm, F. *J. comp. Neurol.* **240**, 143–152 (1985).
6. Nottebohm, F. *Ann. N.Y. Acad. Sci.* **457**, 143–161 (1985).
7. Nottebohm, F. in *Handbook of Physiology, The Nervous System* Vol. 5 (eds Mountcastle, V. & Plum, F.) Part II (American Physiological Society, Washington, 1987).
8. Alvarez-Buylla, A., Theelen, M. & Nottebohm, F. *J. Neurosci.* (in the press).
9. Boulder Committee *Anat. Rec.* **166**, 257–261 (1970).
10. Alvarez-Buylla, A., Buskirk, D. R. & Nottebohm, F. *J. comp. Neurol.* **264**, 159–170 (1987).
11. Jacobson, M. *Developmental Neurobiology*, 66–87 (Plenum, New York, 1978).
12. Edmondson, J. C. & Hatten, M. E. *J. Neurosci.* **7**, 1928–1934 (1987).
13. Rakic, P. *J. comp. Neurol.* **141**, 283–312 (1971).
14. Rakic, P. *Trends Neurosci.* **4**, 184–187 (1981).
15. Rakic, P. in *The Cell in Contact: Adhesions as Morphogenetic Determinants* (eds Edelman, G. E. & Thiery, J. P.) (Wiley, New York, 1986).
16. Stokes, T., Leonard, C. M. & Nottebohm, F. *J. comp. Neurol.* **156**, 337–374 (1974).
17. Alvarez-Buylla, A. & Vicario, D. S. *J. Neurosci. Meth.* (in the press).

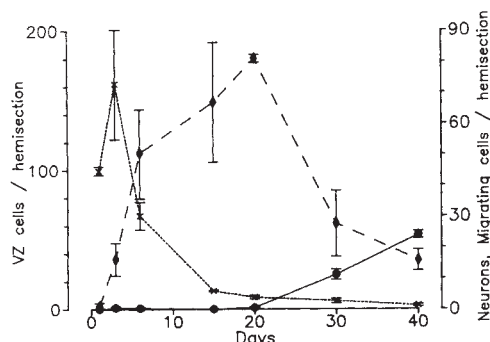


Fig. 3 Number of labelled ventricular zone cells, $\times \cdots \times$ (left-side scale), labelled migrating cells; $\times \cdots \times$ and labelled neurons, $\bullet \cdots \bullet$ (right-side scale) at different survival times after [3 H]thymidine treatment (see Fig. 2 for experimental details). Each point is the mean of results from three birds (four hemisections per bird) $\pm$ standard error.

Selective coupling with K⁺ currents of muscarinic acetylcholine receptor subtypes in NG108-15 cells

Kazuhiko Fukuda, Haruhiro Higashida*, Tai Kubo, Akito Maeda, Isamu Akiba, Hideaki Bujo, Masayoshi Mishina & Shosaku Numa

Departments of Medical Chemistry and Molecular Genetics, Kyoto University Faculty of Medicine, Kyoto 606, Japan

* Department of Biophysics, Neuroinformation Research Institute, Kanazawa University School of Medicine, Kanazawa 920, Japan

The primary structures of two muscarinic acetylcholine receptor (mAChR) species, designated as mAChR I and mAChR II, have been elucidated by cloning and sequence analysis of DNAs complementary to the porcine cerebral and cardiac messenger RNAs, respectively¹⁻³. mAChR I and mAChR II expressed in *Xenopus* oocytes differ from each other both in acetylcholine-induced response and in antagonist binding properties^{1,4}. These results, together with the differential tissue location of the two mAChR mRNAs^{1,2}, have indicated that pharmacologically distinguishable subtypes of the mAChR represent distinct gene products. The primary structures of two additional mammalian mAChR species, designated as mAChR III and mAChR IV, have subsequently been deduced from the nucleotide sequences of the cloned cDNAs or genomic DNAs⁵⁻⁷. We report here that mAChR I and mAChR III expressed in NG108-15 neuroblastoma-glioma hybrid cells, but not mAChR II and mAChR IV, efficiently mediate phosphoinositide hydrolysis, activation of a Ca²⁺-dependent K⁺ current and inhibition of the M-current, a voltage-dependent K⁺ current sensitive to muscarinic agonists.

NG108-15 cells⁸ were transfected with cDNA or genomic DNA encoding each of the four mAChR subtypes, using an expression vector carrying the SV40 early gene promoter and the neomycin-resistance marker gene; mAChR III and mAChR IV are defined as in ref. 5 and correspond to HM4 and HM3 in ref. 6, respectively. Neomycin-resistant clones that showed (–)[³H]quinuclidinyl benzilate (QNB) binding activities more than threefold higher than the activity of non-transfected cells were selected for further studies (Table 1). Blot hybridization analysis of total cellular RNA using mAChR subtype-specific probes (see Table 1 legend) indicated that the transformed clones expressed RNA species (with sizes expected from the expression constructs) encoding the corresponding mAChR subtypes (data not shown). The mAChR IV-specific probe revealed an additional hybridizable RNA species of ~3,200 nucleotides, which was also found in nontransfected or vector-transformed control cells, thus representing the endogenous mAChR IV mRNA in NG108-15 cells⁶. Control cells showed no detectable levels of RNA species hybridizable with any of the three other mAChR subtype-specific probes. Antagonist binding studies of the individual mAChR species, expressed in *Xenopus* oocytes, have indicated that mAChR I, mAChR II and mAChR III correspond most closely to the pharmacologically defined M₁ (I), M₂ cardiac (II) and M₂ glandular (III) subtypes, respectively^{1,4,7}.

The transformed clones were tested for electrophysiological response to focal application of acetylcholine (ACh) under voltage clamp at a holding potential of –30 mV. Cells transformed to express mAChR I or mAChR III exhibited a similar response to ACh. A typical response comprised an initial outward current followed by a sustained inward current (Fig. 1A and Fig. 2A). Both these currents were abolished by atropine (0.1 μM) in a reversible manner. The initial outward current was accompanied by an increase in input conductance (Fig. 1A, B and Fig. 2A, B). The outward current was reversed to an inward current on hyperpolarizing cells to –73 to –85 mV (eight

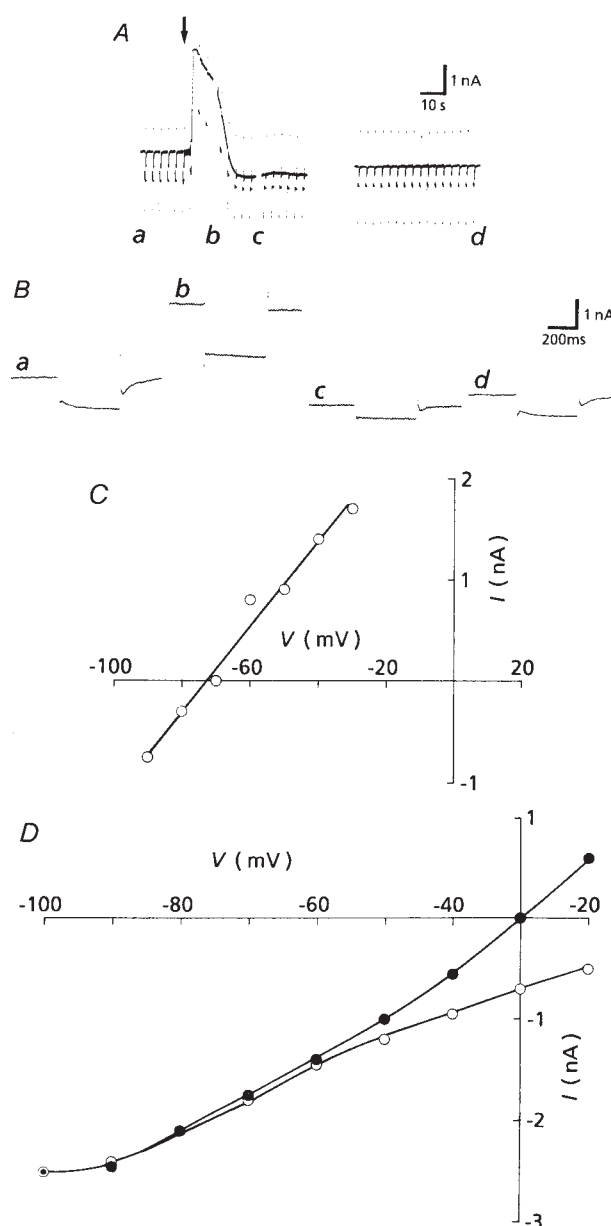


Fig. 1 ACh response in mAChR I-transformed cells (A, B, D, clone NGPM1-3; C, clone NGPM1-8). A, Whole-cell currents activated by applying 3 μl 0.1-mM ACh solution at the time indicated by the arrow were recorded from a voltage-clamped cell by a discontinuous single-electrode voltage-clamp amplifier (Axioclamp-2A, Axon Instruments), holding potential –30 mV, inward current downwards. Repetitive downward deflections are current transients produced by hyperpolarizing steps of 20 mV for 400 ms, applied every 3 s to measure input conductance. An interval of ~5 min separates the trace into two parts and the trace is interrupted by brief periods (~1 s) of faster recording. The recorder saturated during the outward current response. B, Expanded records of current transients obtained at the times indicated in A: a, before applying ACh; b, during the ACh-induced outward current; c, during the subsequent inward current; d, after partial recovery. C, Effect of changing clamp holding potential on the ACh-induced outward current. Whole-cell currents activated by applying 3 μl 0.1-mM ACh solution were recorded under voltage clamp at different holding potentials. Initial peak current activated by ACh is plotted against the holding potential. The reversal potential, –73 mV. D, I–V relationships obtained before adding ACh (●) and during the sustained inward current induced by applying 3 μl 0.1-mM ACh solution (○). The current was measured as the displacement from the original holding current at the end of a series of incrementing 400 ms voltage steps from a holding potential of –30 mV.

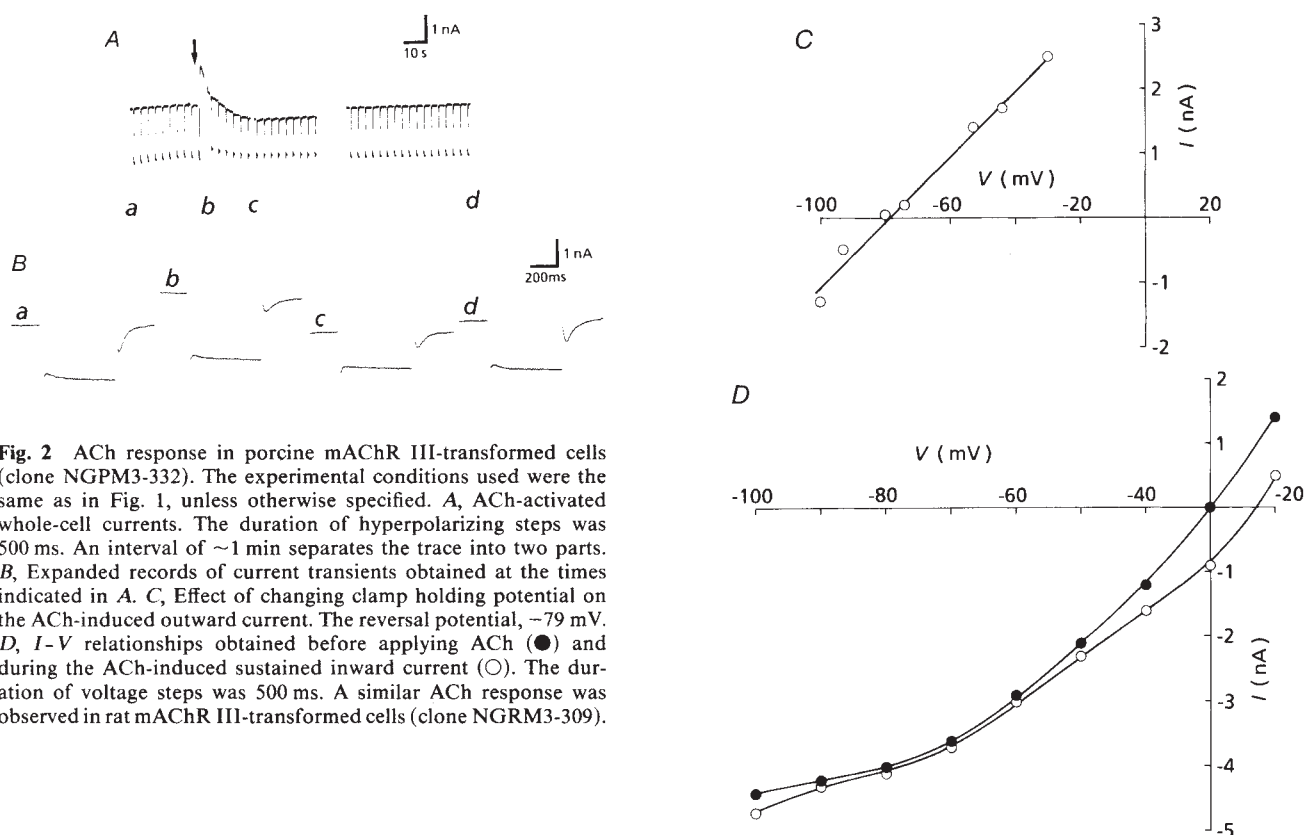
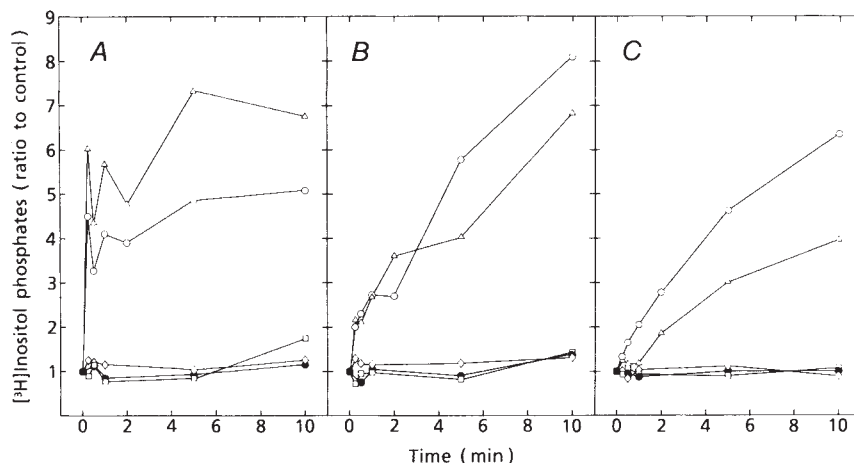


Fig. 2 ACh response in porcine mAChR III-transformed cells (clone NGPM3-332). The experimental conditions used were the same as in Fig. 1, unless otherwise specified. **A**, ACh-activated whole-cell currents. The duration of hyperpolarizing steps was 500 ms. An interval of ~ 1 min separates the trace into two parts. **B**, Expanded records of current transients obtained at the times indicated in **A**. **C**, Effect of changing clamp holding potential on the ACh-induced outward current. The reversal potential, ~ 79 mV. **D**, I - V relationships obtained before applying ACh ($\bullet$) and during the ACh-induced sustained inward current ($\circ$). The duration of voltage steps was 500 ms. A similar ACh response was observed in rat mAChR III-transformed cells (clone NGRM3-309).

Fig. 3 Effect of carbachol on the production of **A**, [3 H]inositol trisphosphate; **B**, [3 H]inositol biphosphate and **C**, [3 H]inositol monophosphate in nontransfected NG108-15 cells ($\bullet$) and transfected cells with porcine mAChR I (NGPM1-27) ($\circ$), porcine mAChR II (NGPM2-106) ($\square$), porcine mAChR III (NGPM3-332) ($\triangle$) or rat mAChR IV (NGRM4-205) ($\diamond$). [3 H]Inositol-labelled cells were incubated at 37°C for the indicated times with or without 1 mM carbachol. The individual inositol phosphates were separated by anion-exchange chromatography²² and counted for radioactivity. Data are expressed as ratios to the respective controls. The values are means of two experiments ($\circ, \square, \triangle$) or from single experiments ($\bullet, \diamond$).



experiments) and the current-voltage (I - V) relation was apparently linear (Fig. 1C and Fig. 2C). The outward current response was reduced or abolished by apamin ($0.4 \mu\text{M}$) or (+)tubocurarine (0.2 mM), but was resistant to tetraethylammonium (1 mM). These results indicate that the initial outward current is principally attributable to activation of a subclass of Ca^{2+} -dependent K^+ currents analogous to I_{AHP} (ref. 9).

The secondary inward current induced by ACh was usually accompanied by a decrease in input conductance and by a reduction in the time-dependent inward current relaxation observed during hyperpolarizing voltage steps (Fig. 1A, B and Fig. 2A, B). The steady-state I - V curve that was obtained while the sustained inward current was present showed a reduction in the outward rectification in a potential range less negative than about -70 mV (tested up to -20 mV), where the M-current^{10,11} (I_{M}) is activated (Fig. 1D and Fig. 2D). These data indicate that the secondary inward current results primarily from inhibition of I_{M} .

As shown in Table 1, the initial outward current (10–30 s in duration) was observed in 58 of 66 mAChR I-transformed cells

and in 87 of 108 mAChR III-transformed cells. The secondary inward current (longer than 2 min in duration) occurred in 39 of 66 mAChR I-transformed cells and in 69 of 108 mAChR III-transformed cells. In contrast, the percentages of responsive cells among mAChR II- or mAChR IV-transformed cells, despite their ($-$)[^3H]QNB binding activities being comparable to those of mAChR I- or mAChR III-transformed cells, were not appreciably higher than those of nontransfected or vector-transformed control cells. The ACh responses observed in mAChR II- or mAChR IV-transformed cells and in control cells were not further characterized.

The effect of carbachol on the formation of [^3H]inositol phosphates was next examined, using [^3H]inositol-labelled cells. In mAChR I- or mAChR III-transformed cells, a four- to seven-fold increase in the release of total inositol phosphates occurred in response to carbachol stimulation, as compared with control values obtained without stimulation (Table 1). But in mAChR II- or mAChR IV-transformed cells, as well as in control cells, no appreciable increase in inositol phosphate release was observed. In mAChR I- or mAChR III-transformed cells, the release of

Table 1 Functional properties of transformed NG108-15 cells

Transformed with	Clone	(-)[³ H]QNB binding (fmol per mg protein)	ACh-activated current				Carbachol-stimulated [³ H]inositol phosphate formation (ratio to control)
			Outward	Number of responsive cells/number tested	Inward	Number of responsive cells/number tested	
			Mean amplitude (nA)		Mean amplitude (nA)		
None	NG108-15	107 ± 31 (23)	4/22	0.2 ± 0.1	2/22	0.3	1.0 ± 0.1 (5)
Vector	NGV-101	142 ± 49 (4)	1/16	0.1	0/16	—	1.2 ± 0.2 (4)
	NGV-102	103 ± 34 (3)	4/28	0.5 ± 0.4	2/28	0.2	1.1 ± 0.1 (3)
	NGPM1-3	812 ± 173 (5)	9/10	0.8 ± 1.1	4/10	0.3 ± 0.3	7.2 ± 2.6 (3)
mAChR I (porcine)	NGPM1-8	831 ± 273 (6)	21/22	1.7 ± 1.3	10/22	0.8 ± 0.8	6.4 ± 2.1 (3)
	NGPM1-27	1,182 ± 148 (4)	28/34	2.1 ± 1.3	25/34	0.7 ± 0.5	7.4 ± 1.4 (3)
	NGPM2-102	656 ± 20 (3)	3/24	0.8 ± 0.5	3/24	0.2 ± 0.1	1.2 ± 0.2 (4)
mAChR II (porcine)	NGPM2-105	611 ± 204 (3)	2/25	0.5	2/25	0.5	1.1 ± 0.1 (3)
	NGPM2-106	638 ± 108 (5)	4/21	0.9 ± 1.2	0/21	—	1.1 ± 0.1 (3)
	NGPM3-161	424 ± 28 (3)	21/32	1.5 ± 1.6	15/32	0.4 ± 0.3	6.9 ± 1.6 (3)
mAChR III (porcine)	NGPM3-332	842 ± 95 (3)	41/44	1.6 ± 1.2	38/44	0.5 ± 0.4	6.0 ± 2.2 (3)
	NGRM3-309	369 ± 13 (3)	12/18	2.6 ± 1.7	9/18	0.4 ± 0.3	4.3 ± 0.6 (3)
mAChR III (rat)	NGRM3-310	359 ± 57 (4)	13/14	1.1 ± 1.1	7/14	0.4 ± 0.4	4.4 ± 0.2 (3)
	NGRM4-205	390 ± 128 (3)	1/15	1.2	0/15	—	0.9 ± 0.1 (3)

Data are given as means ± s.d. when indicated (number of experiments in parentheses). Whole-cell currents activated by 0.1 mM or 1 mM ACh (3 μ l) applied close to the cell surface were recorded under voltage clamp at -30 mV holding potential; the means were calculated only for the responsive cells that showed ACh-activated currents larger than ~0.1 nA. The formation of total [³H]inositol phosphates in cells stimulated by incubation with 1 mM carbachol at 37 °C for 10 min is expressed as the ratio to that in control cells incubated without carbachol. The 8.2-kilobase-pair (kb) *Hind*III fragment from the plasmid pKGS α N (ref. 18) was circularized to generate pKNH. pKNS and pKNP are derivatives of pKNH carrying a *Sac*I site and a polylinker sequence, respectively, in place of the *Hind*III site. Using synthetic linkers, mAChR cDNAs or genomic DNAs (containing no introns in the protein-coding region) were inserted into an expression plasmid as follows: pKPM1 (porcine mAChR I), *Eco*RI/*Bst*EII fragment from nucleotide -171 to 1,495 (ref. 1) into the *Hind*III site of pKNH; pKPM2 (porcine mAChR II), *Hind*III/*Sac*I fragment from nucleotide -92 to 1,467 (ref. 2) into the *Hind*III site of pKNH; pKPM3 (clone NGPM3-332) and pKPM3X (clone NGPM3-161) (porcine mAChR III), 2.1-kb *Tth*1111/*Sma*I fragment including the 16-base-pair (bp) 5'-noncoding and the ~340-bp 3'-noncoding sequence⁷ into the *Sac*I site of pKNS and into the *Xba*I site of pKNP, respectively; pKRM3 (rat mAChR III), 1.8-kb *Tth*1111/*Fnu*DII fragment⁷ into the *Sac*I site of pKNS; pKRM4 (rat mAChR IV), 1.6-kb segment including the 11-bp 5'-noncoding and the ~180-bp 3'-noncoding sequence up to the *Sal*I site⁷ into the *Hind*III site of pKNH. NG108-15 cells in culture¹⁹ were transfected²⁰ with the above plasmids and screened for G418-resistant clones. (-)[³H]QNB binding activity was determined as described previously¹³, except that membrane preparations (~100 μ g protein), obtained by centrifugation at 20,000g for 30 min, were incubated with 1 nM (-)[³H]QNB (43.6 Ci mmol⁻¹) at 30 °C for 1 h. Electrophysiological measurements¹¹ and the assay of inositol phosphate formation^{21,22} were carried out essentially as described previously. The mAChR subtype-specific probes used for RNA blot hybridization analysis²³ were restriction fragments encoding the following amino acid residues (in parentheses): porcine mAChR I (226-343); porcine mAChR II (237-380); porcine mAChR III (262-463); rat mAChR III (270-460); rat mAChR IV (264-392).

inositol triphosphate increased rapidly (Fig. 3A). This was followed by an increase in the formation of inositol bisphosphate (Fig. 3B) and inositol monophosphate (Fig. 3C). These results suggest that mAChR I and mAChR III mediate stimulation of phospholipase C, thus enhancing release of inositol 1,4,5-trisphosphate, which is known to mobilize Ca²⁺ from intracellular stores¹². The formation of inositol phosphates and the biphasic changes in membrane conductance induced by agonist in mAChR I- or mAChR III-transformed cells were not appreciably affected by prior incubation of the cells with pertussis toxin (100 ng ml⁻¹; 24 h). The inhibition by agonist of prostaglandin E₁-induced cyclic AMP accumulation observed in these cells and in nontransfected cells¹³ was strongly reduced by the same treatment however (data not shown). This suggests that the responses mediated by mAChR I or mAChR III and the inhibition of adenylate cyclase mediated by the endogenous mAChR (probably the mAChR IV subtype; see above) occur through distinct G-proteins.

The present investigation indicates that the molecularly defined mAChR I and mAChR III subtypes, in contrast to the mAChR II and mAChR IV subtypes, efficiently mediate phosphoinositide hydrolysis, activation of a Ca²⁺-dependent K⁺ current and inhibition of I_M in NG108-15 cells. A similar current response has been shown to be induced by bradykinin through phosphoinositide hydrolysis in NG108-15 cells^{11,14}. It is suggested that mAChR IV mediates adenylate cyclase inhibition without affecting phosphoinositide hydrolysis in NG108-15 cells (see above and ref. 15). Recent work reports that mAChR II expressed in Chinese hamster ovary cells is coupled with adeny-

late cyclase inhibition more efficiently than with phosphoinositide hydrolysis¹⁶. Our previous cDNA expression studies using *Xenopus* oocytes provide evidence that mAChR I induces activation of a Ca²⁺-dependent Cl⁻ current, whereas mAChR II principally evokes activation of Na⁺ and K⁺ currents in a Ca²⁺-independent manner⁴. All these findings suggest that mAChR subtypes are selectively coupled with different effector systems, albeit not exclusively. The inhibition of I_M mediated by mAChR I and mAChR III suggests that these subtypes are involved in the cholinergic excitation of vertebrate neurons evoked by muscarinic activity.

Since the submission of this manuscript, we have learned that human mAChR I (HM1) and mAChR III (HM4) expressed in human kidney cells strongly activate phosphoinositide hydrolysis and that mAChR II (HM2) and mAChR IV (HM3) efficiently inhibit adenylate cyclase¹⁷.

This investigation was supported in part by research grants from the Ministry of Education, Science and Culture of Japan, the Institute of Physical and Chemical Research, the Mitsubishi Foundation and the Japanese Foundation of Metabolism and Diseases.

Received 23 May; accepted 3 August 1988.

1. Kubo, T. *et al.* *Nature* **323**, 411-416 (1986).
2. Kubo, T. *et al.* *FEBS Lett.* **209**, 367-372 (1986).
3. Peralta, E. G. *et al.* *Science* **236**, 600-605 (1987).
4. Fukuda, K. *et al.* *Nature* **327**, 623-625 (1987).
5. Bonner, T. I., Buckley, N. J., Young, A. C. & Brann, M. R. *Science* **237**, 527-532 (1987).
6. Peralta, E. G. *et al.* *EMBO J.* **6**, 3923-3929 (1987).
7. Akiba, I. *et al.* *FEBS Lett.* **235**, 257-261 (1988).
8. Nirenberg, M. *et al.* *Science* **222**, 794-799 (1983).

9. Pennefather, P., Lancaster, B., Adams, P. R. & Nicoll, R. A. *Proc. natn. Acad. Sci. U.S.A.* **82**, 3040-3044 (1985).
10. Brown, D. A. & Adams, P. R. *Nature* **283**, 673-676 (1980).
11. Higashida, H. & Brown, D. A. *Nature* **323**, 333-335 (1986).
12. Berridge, M. J. *A. Rev. Biochem.* **56**, 159-193 (1987).
13. Kurose, H., Katada, T., Amano, T. & Ui, M. *J. biol. Chem.* **258**, 4870-4875 (1983).
14. Yano, K., Higashida, H., Hattori, H. & Nozawa, Y. *FEBS Lett.* **181**, 403-406 (1985).
15. Harden, T. K. *et al. Trends pharmac. Sci. Suppl.* **7**, 14-18 (1986).
16. Ashkenazi, A. *et al. Science* **238**, 672-675 (1988).
17. Peralta, E. G., Ashkenazi, A., Winslow, J. W., Ramachandran, J. & Capon, D. J. *Nature* **334**, 434-437 (1988).
18. Nukada, T., Mishina, M. & Numa, S. *FEBS Lett.* **211**, 5-9 (1987).
19. Higashida, H. *et al. Brain Res.* **214**, 287-299 (1981).
20. Gorman, C. in *DNA Cloning vol. II* (ed. Glover, D. M.) 143-190 (IRL, Oxford, 1985).
21. Bone, E. A., Fretten, P., Palmer, S., Kirk, C. J. & Mitchell, R. H. *Biochem. J.* **221**, 803-811 (1984).
22. Berridge, M. J., Dawson, R. M. C., Downes, C. P., Heslop, J. P. & Irvine, R. F. *Biochem. J.* **212**, 473-482 (1983).
23. Thomas, P. S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5201-5205 (1980).

The genomic clone G-21 which resembles a β -adrenergic receptor sequence encodes the 5-HT_{1A} receptor

Annick Fargin*, John R. Raymond*†, Martin J. Lohse*, Brian K. Kobilka*, Marc G. Caron*‡ & Robert J. Lefkowitz*

Howard Hughes Medical Institute Laboratories: *Departments of Medicine (Cardiology) and Biochemistry, †Medicine (Nephrology) and ‡Cell Biology, Duke University Medical Center, Durham, North Carolina 27710, USA

The recent cloning of the complementary DNAs and/or genes for several receptors linked to guanine nucleotide regulatory proteins including the adrenergic receptors (α_1 , α_{2A} , α_{2B} , β_1 , β_2)¹⁻⁷, several subtypes of the muscarinic cholinergic receptors^{8,9}, and the visual 'receptor' rhodopsin¹⁰ has revealed considerable similarity in the primary structure of these proteins. In addition, all of these proteins contain seven putative transmembrane α -helices. We have previously described a genomic clone, G-21, isolated by cross-hybridization at reduced stringency with a full length β_2 -adrenergic receptor probe¹¹. This clone contains an intronless gene which, because of its striking sequence resemblance to the adrenergic receptors, is presumed to encode a G-protein-coupled receptor. Previous attempts to identify this putative receptor by expression studies have failed. We now report that the protein product of the genomic clone, G21, transiently expressed in monkey kidney cells has all the typical ligand-binding characteristics of the 5-hydroxytryptamine (5-HT_{1A}) receptor.

At least six subtypes of 5-HT receptors (1A, 1B, 1C, 1D, 2 and 3) have been extensively characterized by pharmacological and physiological methods (see refs 12-15 for reviews). These various receptors are coupled to the major transduction systems known to mediate the action of G-protein-coupled receptors. However, nothing is known about the primary structures of any of the 5-HT receptor subtypes, as yet (see note added in proof).

Given the homology between the G-21 and the β -adrenergic receptor genes, membranes from monkey kidney (COS-7) cells transfected with G-21 DNA in the vector pBC12B1¹⁶ were tested for the binding of the β -adrenergic antagonist radioligand [¹²⁵I]-cyanopindolol (¹²⁵I-CYP). Although the membranes displayed specific binding of ¹²⁵I-CYP, saturation experiments revealed only a moderate affinity (~10 nM), more consistent with a 5-HT_{1A} or 5-HT_{1B} binding site^{17,18}. In fact, 5-HT competed more potently for the binding than adrenergic ligands. Subsequent studies with the 5-HT_{1A} selective agonist radioligand [³H]8-OH-2-(di-*n*-propylamino) 1,2,3,4-tetrahydronaphthalene ([³H]8-OH-DPAT)¹⁹ showed saturable specific binding. Two classes of sites with affinities of 0.06 and 14.5 nM were observed (Fig. 1, inset), which is consistent with 8-OH-DPAT being an agonist whose binding may be associated with two affinity states of the

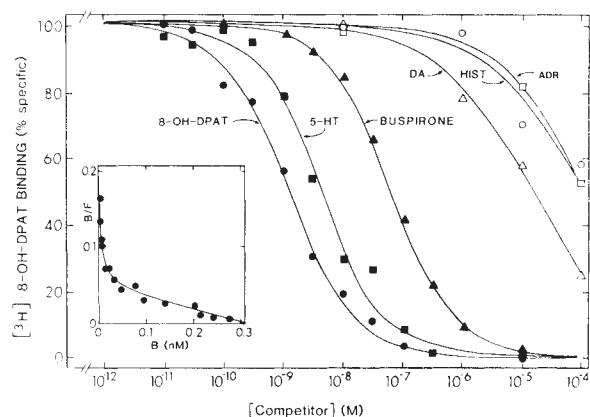


Fig. 1 Binding properties of [³H]8-OH-DPAT to the genomic clone G-21 expressed in COS-7 cells. The ability of various agonists to compete for the binding of [³H]8-OH-DPAT (0.5 nM, 110 Ci mmol⁻¹) in membranes of COS-7 cells transfected with the clone G-21 is depicted. The inset shows the Scatchard plot of saturation binding data for [³H]8-OH-DPAT. Maximal binding (B_{max}) was ~0.5 pmol per mg protein for the high affinity state and ~19 pmol per mg protein for the low affinity state. DA, dopamine; Hist, histamine; ADR, adrenaline.

Methods. Expression vector pSPXB¹¹ containing the *Xba*I-*Bam*HI fragment of the genomic clone G-21 was digested with *Hind*III (in the polylinker of pSP64) and *Bam*HI. The 1.7 kb *Hind*III-*Bam*HI fragment was isolated and inserted between unique *Hind*III and *Bam*HI sites of the expression vector pBC12B1¹⁶. The DEAE-dextran method was used for the transfection of COS-7 cells. At 72 hours after transfection, cells were lysed in ice-cold 20 mM Tris HCl (pH 7.4), 5 mM EDTA. Nuclear debris and intact cells were removed by centrifugation at 1,000g for 10 min. The supernatant was pelleted at 35,000g for 20 min, the resulting pellet washed once in the above described lysis buffer and then resuspended in binding buffer—50 mM Tris HCl (pH 7.4), 2.5 mM MgCl₂, 10 μ M pargyline—at a final protein concentration of ~160 μ g ml⁻¹. Incubation (30 min, 30 °C) with the appropriate concentration of competing ligands was terminated by the addition of 5 ml of ice-cold binding buffer, rapid vacuum filtration through Whatman GF/C filters, and two subsequent 5 ml washes. For the saturation binding data (inset) nonspecific binding was determined with 10 μ M 5-HT. Typical values of total and nonspecific binding at a radioligand concentration of 0.5 nM were 1,900 c.p.m. and 200 c.p.m., respectively. Data are mean values of triplicate determinations from representative experiments. Four separate transfections gave an expression of 14 $\pm$ 2 pmol per mg membrane protein. Mean K_i values of competing ligands are given in Table 1.

receptor²⁰⁻²². Figure 1 shows competition curves of various agonists with [³H]8-OH-DPAT. At the radioligand concentration used for these experiments (0.5 nM) more than 90% of the binding was competed for by 1 μ M 5-HT. Among the agonists, the order of potency was 8-OH-DPAT > ipsapirone ~ 5-HT > buspirone > dopamine > adrenaline = histamine. Among antagonists, the order of potency was spiroxatrine > spiperone > mesulergine > ketanserin. The calculated dissociation constants of the various drugs are summarized in Table 1. The data unequivocally identify the binding site as that of the 5-HT_{1A} receptor.

The nature of the effector system coupled to the 5-HT_{1A} receptor has been controversial, with various reports suggesting stimulation^{23,24} or inhibition²⁵⁻²⁷ of adenylate cyclase, or stimulation of a K⁺-conductance^{28,29}. Regardless of which system(s) are involved, it is likely that a G-protein mediates receptor coupling, in which case guanine nucleotides would be able to decrease high-affinity agonist binding to the receptors by converting them to a low affinity state. Accordingly, we tested the ability of several guanine nucleotides to decrease binding of

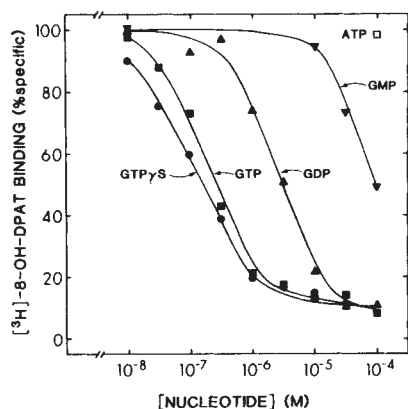


Fig. 2 Effects of guanine nucleotides on [^3H]8-OH-DPAT binding to membranes from G-21 transfected COS-7 cells. Membranes were resuspended in Mg^{2+} -free binding buffer at a protein concentration of $320 \mu\text{g ml}^{-1}$ and were incubated (30 mins, 30°C) with 0.1 nM [^3H]8-OH-DPAT in the presence of various concentrations of nucleotides. 50% inhibitory concentration (IC_{50})-values were $0.1 \mu\text{M}$ for GTP γS , $0.2 \mu\text{M}$ for GTP, $2 \mu\text{M}$ for GDP, and $70 \mu\text{M}$ for GMP (means of two experiments).

[^3H]8-OH-DPAT 0.1 nM to the receptors. As shown in Fig. 2, GTP γS (guanosine 5'-O-(thiotriphosphate)) and GTP had a potent effect: GDP was weaker, GMP much weaker still and ATP was inert. These results are in good agreement with those reported for [^3H]8-OH-DPAT binding to 5-HT $_{1A}$ receptors in rat brain membranes 21,22 .

Several additional lines of evidence further document the relationship between expression of G-21 and appearance of the 5-HT $_{1A}$ receptor binding activity. First, transfection without the plasmid, with the plasmid lacking the G-21 DNA, or with plasmid containing either of two separate α_2 -adrenergic receptor genes (α_2A and α_2B , refs 5 and 7) did not yield specific binding under the same conditions. Second, Northern blots were performed with RNA prepared from transfected and non-transfected COS-7 cells. A single hybridizing mRNA species of the expected size ($\sim 2.3 \text{ kb}$) was observed in the transfected but not in the non-transfected cells or in cells transfected with the plasmid only (Fig. 3).

Table 1 Affinities of ligands for [^3H]8-OH-DPAT binding sites

Agonists	Compound	K_H (nM)	K_L (nM)
	8-OH-DPAT	0.06 ± 0.01	22 ± 6.4
	Ipsapirone	0.24 ± 0.08	31 ± 2.7
	5-HT	0.27 ± 0.07	63 ± 5.5
	Buspirone	4.0 ± 1.6	97 ± 25
	Dopamine	$>1,000$	ND
	Histamine	$>5,000$	ND
	Adrenaline	$>5,000$	ND
Antagonists	Compound	K_i (nM)	
	Spiroxtarine	1.8 ± 0.7	
	Spiperone	63 ± 28	
	Mesulergine	400 ± 62	
	Ketanserin	$2,100 \pm 450$	

Membranes from G-21-transfected COS-7 cells were used. The ability of various compounds to compete for the specific binding of [^3H]8-OH-DPAT is shown. Results of competition curves were analysed for two affinity states for agonists (K_H and K_L) and one for antagonists (K_i) in agreement with previous studies of 5-HT $_{1A}$ receptors $^{20-22}$ as described 33 . Values represent the means and standard error of the means of at least three experiments. ND, not determined.

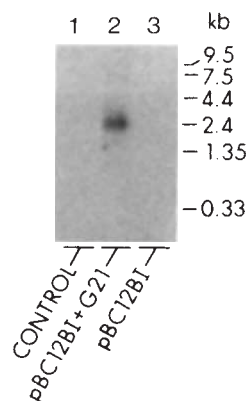


Fig. 3 Northern blot analysis of G-21 transcripts in COS-7 cells. Lane 1 contains RNA from non-transfected control cells, lane 2, RNA from cells transfected with G-21 DNA in the vector pBC12BI, and lane 3, RNA from cells transfected with the vector alone. The size of the message obtained ($\sim 2.3 \text{ kb}$) is consistent with the size expected for the expression of the *HindIII*-*BamHI* fragment of G-21 in the pBC12BI vector. This message size is different from that of the natural message (6.0 kb) of G-21 11 .

Methods. Total RNA was prepared from the cells as described 34 . Samples ($5 \mu\text{g}$) were denatured in glyoxal 35 , electrophoresed on 1.2% agarose gels and transferred for blotting 36 . The filter was hybridized at 42°C for 24 hours in a cocktail containing 50% formamide, $5 \times \text{SSC}$, $1 \times \text{Denhardt's}$ SDS (sodium dodecyl sulphate) 10% dextran sulphate, and $100 \mu\text{g ml}^{-1}$ salmon sperm DNA, using a full length G-21 probe ($4 \times 10^6 \text{ d.p.m. ml}^{-1}$). Final washing conditions were $0.1 \times \text{SSC}$, 0.1% SDS, 60°C .

To confirm that the protein encoded by the G-21 gene in fact represents the 5-HT $_{1A}$ receptor binding site we used anti-peptide antibodies to the peptide corresponding to amino acids 242–267, present in the putative third cytoplasmic loop of the G-21 protein. The antibodies immunoprecipitated [^3H]8-OH-DPAT specifically bound to soluble receptors, and this was completely blocked by the corresponding synthetic peptide (Fig. 4). Comparable anti-peptide antibodies raised to the β_2 -adrenergic receptor did not precipitate any radioactivity above the blank nor did a peptide sequence from elsewhere in the G-21 protein (amino acids 268–293) block the precipitation by the anti-G-21 antibody (data not shown).

The identification of the protein product of the G-21 clone as the 5-HT $_{1A}$ receptor raises several interesting issues. The remarkable sequence similarity between G-21 and all of the adrenergic receptors ($\sim 45\%$ of residues in the presumed membrane-spanning domains are identical between G-21 and any adrenergic receptor) had earlier led us to speculate that G-21 encoded an adrenergic receptor 11 . Clearly, however, its protein product represents a receptor not for adrenaline, but for another biogenic amine, 5-HT. Moreover, the 5-HT $_{1A}$ and 5-HT $_{1B}$ receptors are further distinguished from other 5-HT receptor subtypes by their ability to bind β -adrenergic antagonists such as cyanopindolol and propranolol (but not agonists) with moderate affinity 15,17,18 . Using a series of chimaeric receptor genes constructed from the genes for α_2 - and β_2 -adrenergic receptors, we recently demonstrated that the 6th and 7th membrane-spanning domains of these receptors are very important in determining their antagonist binding properties 30 . G-21 is most similar to the β_2 -adrenergic receptor in the 6th membrane-spanning domain (20/26 residues are identical), suggesting that these structure-function relationships may extend to a variety of G-protein-coupled receptors.

We have previously shown that the G-21 mRNA is abundant in a variety of lymphoid tissues 11 . This emphasizes the potential importance of 5-HT as an immunomodulator 31,32 . The identification of clone G-21 as the 5-HT $_{1A}$ receptor further

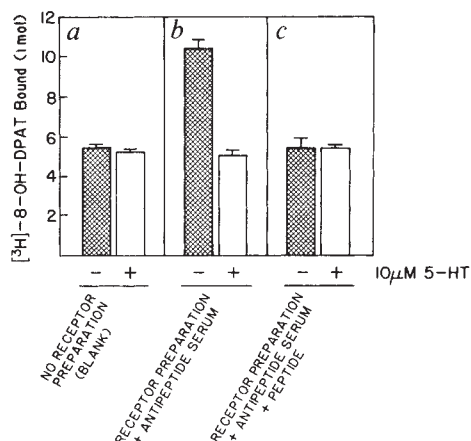


Fig. 4 Immunoprecipitation of [3 H]8-OH-DPAT-binding activity of the protein product of G-21 expressed in COS-7 cells. COS-7 cell membranes incubated with [3 H]8-OH-DPAT in the presence or absence of 10 μ M 5-HT were solubilized and immunoprecipitated with an antiserum to a peptide deduced from the DNA sequence of G-21. The immunoprecipitation procedure was carried out with the G-21 peptide antiserum: *a*, in the absence of COS-7 cell soluble preparation; and *b*, in the presence of solubilized membranes. *c*, Immunoprecipitation was carried out as in *b* except the antiserum was preincubated with the corresponding peptide (10 μ M). The data shown represent the means of two separate experiments performed in duplicate; the bars indicate the s.e.m.

Methods. Immunoprecipitation: COS-7 cell membrane from cells transfected with the G-21 plasmid prepared as in Fig. 1 were incubated with 10 nM [3 H]8-OH-DPAT in the presence or absence of 10 μ M 5-HT for 30 min at 22 $^{\circ}$ C. Membranes were then solubilized in binding buffer including 10 mM CHAPS (3-[(3-cholamidopropyl)-dimethylammonio]-1-propane-sulphonate), 200 mM NaCl³⁷ with the addition of 2 mM EDTA and 5 μ g ml⁻¹ each of leupeptin and benzamidin for 1 hour at 0 $^{\circ}$ C, and centrifuged at 500,000g for 30 min. The concentrations of 5-HT and [3 H]8-OH-DPAT were maintained throughout the experiment. The supernatant was diluted 8-fold with binding buffer including the protease inhibitors and 10 mM ascorbate, and incubated with antiserum at a 1:25 dilution in an assay volume of 250 μ l. Under some conditions, the G-21 peptide antiserum was preincubated with 10 μ M of the corresponding or non-corresponding peptide for 2 h at 22 $^{\circ}$ C. The immune complexes were immobilized by adding 15 μ l of 50% (w/v) pre-swollen protein A-Sepharose 6MB to each assay and shaking for 1 h at 4 $^{\circ}$ C. Specific [3 H]8-OH-DPAT binding to receptor-antibody-protein A-Sepharose 6MB-complexes was measured after terminating the assay with 5 ml of ice-cold binding buffer (without [3 H]8-OH-DPAT or 5-HT), followed by rapid vacuum filtration followed by two rinses over glass fibre (GF/F) filters. **Antisera:** polyclonal antisera were raised in rabbits to a sequence of the putative third intracellular loop of the G-21 protein product (amino acids 242-267: GASPAPQPKKS-VNGESGSRNWRLGVE) coupled to keyhole limpet haemocyanin. Antisera were screened for their ability to recognize the corresponding peptides by a solid phase radioimmunoassay and for their ability to immunoblot the expressed G-21 protein product in transfected versus non-transfected COS-7 cell membranes (J. R. Raymond *et al.*, in preparation). A peptide from another portion of the G-21 protein (amino acids 268-293: SKAG-GALCANGAVRFGGALELVIE) was used in these experiments to establish specificity.

demonstrates the remarkable evolutionary relationships between the various members of the extended family of G-protein-coupled receptors. Moreover, its expression in various cell lines will be extremely useful in the clarification of the physiological roles and biochemical signalling mechanisms of the 5-HT_{1A} receptor.

We thank Drs J. E. Leysen, M. R. Hnatowich, and J. M. Palacios for helpful discussion and suggestions and Dr J. W. Regan for providing the synthetic peptides and the COS-7 cells

transfected with human α_2A and α_2B_A receptor genes. A.F. was supported by a fellowship from the Ministère des Affaires Étrangères, the Fondation pour la Recherche Médicale, and the Association pour la Recherche sur le Cancer of France. M.J.L. was supported by a fellowship from the Deutsche Forschungsgemeinschaft.

Note added in proof: Since submission of this manuscript, a report on the sequence of the 5-HT_{1C} receptor has appeared³⁸.

Received 18 July; accepted 5 August 1988.

- Dixon, R. A. F. *et al.* *Nature* **321**, 75-79 (1986).
- Yarden, Y. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **83**, 6795-6799 (1986).
- Kobilka, B. K. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **84**, 46-50 (1987).
- Frielle, T. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **84**, 7920-7924 (1987).
- Kobilka, B. K. *et al.* *Science* **238**, 650-656 (1987).
- Cotecchia S. *et al.* *Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Regan, J. W. *et al.* *Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Kubo, T. *et al.* *Nature* **323**, 411-416 (1986).
- Kubo, T. *et al.* *FEBS Lett.* **209**, 367-372 (1986).
- Nathans, J. & Hogness, D. S. *Proc. natn. Acad. Sci. U.S.A.* **81**, 4851-4855 (1984).
- Kobilka, B. K. *et al.* *Nature* **329**, 75-79 (1987).
- Bradley, P. B. *et al.* *Neuropharmacol.* **25**, 563-575 (1987).
- Göthert, M. & Schlicker, E. *J. Cardiovasc. Pharmacol.* **10**(suppl. 10), S3-S7 (1987).
- Peroutka, S. J. *ISI Atlas of Sci.: Pharmac.* **2**, 1-4 (1988).
- Leysen, J. E. *Neuromethods, Neuropharmacology II: Drugs as Tools in Neurotransmitter Research* (eds Boulton, A. A., Baker, G. B. & Jourio, A. V.) (Humana, Clifton, in press).
- Cullen, B. R. *Meth. Enzym.* **152**, 684-704 (1987).
- Hoyer, D., Engel, G. & Kalkman, H. O. *Eur. J. Pharmac.* **118**, 1-12 (1985).
- Hoyer, D., Engel, G. & Kalkman, H. O. *Eur. J. Pharmac.* **118**, 13-23 (1985).
- Gozlan, H., Mestikawy, S., Pichat, L., Glowinsky, G. & Hamon, M. *Nature* **305**, 140-142 (1983).
- Sills, M. A., Wolfe, B. B. & Frazer, A. *Molec. Pharmac.* **26**, 10-18 (1984).
- Hall, M. D. *et al.* *J. Neurochem.* **44**, 1685-1696 (1985).
- Schlegel, J. R. & Peroutka, S. J. *Biochem. Pharmac.* **35**, 1943-1949 (1986).
- Shenker, A. *et al.* *Eur. J. Pharmac.* **109**, 427-429 (1985).
- Markstein, R., Hoyer, D. & Engel, G. *Naunyn-Schmiedeberg's Archs Pharmac.* **333**, 335-341 (1986).
- Weiss, S., Sebben, M., Kempe, D. & Bockaert, J. *Eur. J. Pharmacol.* **120**, 227-230 (1986).
- De Vivo, M. & Maayani, S. *J. Pharmac. exp. Ther.* **238**, 248-253 (1986).
- Bockaert, J. A. *et al.* *Naunyn-Schmiedeberg's Arch. Pharmac.* **335**, 588-592 (1987).
- Colino, A. & Halliwell, J. V. *Nature* **328**, 73-77 (1987).
- Andrade, R. & Nicoll, R. A. *J. Physiol., Lond.* **394**, 99-124 (1987).
- Kobilka, B. K. *et al.* *Science* **240**, 1310-1316 (1988).
- Sternberg, E. M., Wedner, J. H., Leung, M. K. & Parker, C. W. *J. Immun.* **138**, 4360-4365 (1987).
- Hellstrand, K. & Hermodsson, S. *J. Immun.* **139**, 869-875 (1987).
- DeLean, A., Hancock, A. A. & Lefkowitz, R. J. *Molec. Pharmac.* **21**, 5-16 (1982).
- Chirgwin, J. M., Przybyla, A. E., McDonald, R. J. & Rutter, W. J. *Biochemistry* **18**, 5294-5299 (1979).
- McMaster, G. K. & Carmichael, G. C. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4835-4838 (1977).
- Azis, N. & Munro, H. N. *Nucleic Acids Res.* **14**, 915-927 (1986).
- Gozlan, H. *et al.* *J. Receptor Res.* **7**, 195-221 (1987).
- Julius, D. *et al.* *Science* **241**, 558-564 (1988).

Membrane potential has no direct role in evoking neurotransmitter release

Robert S. Zucker* & Philip G. Haydon†

*Department of Physiology-Anatomy, University of California, Berkeley, California 94720, USA

†Department of Zoology, Iowa State University, Ames, Iowa 50011, USA

Neurons communicate by secreting a transmitter that excites or inhibits other neurons at synapses. The role of presynaptic membrane potential in triggering transmitter release is still controversial. In one view, presynaptic action potentials trigger the release by the entry of calcium ions into presynaptic terminals through voltage-dependent calcium channels¹. Calcium acts at high local concentrations at release sites near channel mouths to cause neurosecretion²⁻⁵. An opposing view is that, in addition to elevating presynaptic calcium, presynaptic potential stimulates transmitter release by a distinct direct action^{1,6-15}. The relative importance of depolarization and calcium entry in neurosecretion cannot be determined because the two events are tightly linked¹⁶⁻²⁰. To delineate the roles of presynaptic potential and calcium entry in transmitter release, we have used nitr-5, a photolabile calcium chelator²¹⁻²², and a voltage-clamp technique to control intracellular calcium and membrane potential independently at a synapse

formed between cell bodies of cultured neurons of the fresh water snail *Helisoma trivolvis*. We found transmitter release occurred when presynaptic calcium levels were elevated to concentrations of a few micromolar, and that presynaptic voltage had no direct effect on neurosecretion.

Buccal ganglion neuron B5 forms cholinergic chloride-dependent inhibitory synapses onto neuron B19 in cell culture. If neurite outgrowth is suppressed by preventing adherence of the neurons to a substrate, the inhibitory synapse forms between the cell bodies²³. This synapse resembles other immature synapses in many respects^{24–26}; in particular, repetitive stimulation evokes postsynaptic potentials of decreasing amplitude due to synaptic depression.

In these experiments, we monitored inhibitory postsynaptic current (i.p.s.c.) in neuron B19 using the whole-cell patch-clamp technique²⁷. I.p.s.c.s were converted to inward currents and magnified by holding the membrane potential at -80 mV and by perfusing the interior of the postsynaptic cell with chloride ions at high concentrations (50 mM) so that the chloride equilibrium potential approaches zero. After recording spontaneous transmitter release, nitr-5 loaded with calcium was introduced into presynaptic neuron B5 using a second whole-cell patch pipette, and presynaptic membrane potential was controlled by a second patch-clamp circuit.

We examined the roles of voltage and calcium in triggering neurotransmitter release. Figure 1a shows a membrane current record from a voltage-clamped postsynaptic neuron B19 before recording from the presynaptic cell and filling it with nitr-5. Miniature i.p.s.c.s are evident, reflecting the spontaneous release of transmitter quanta. A micropipette containing nitr-5 was sealed to neuron B5 and the cell clamped to a waveform that depolarizes the neuron from -40 mV to $+40$ mV, then hyperpolarizes it to -120 mV, and finally repolarizes it to -40 mV (Fig. 1b). Depolarization of the presynaptic neuron causes the rate of transmitter release to rise because the bathing medium contains calcium. Figure 1d shows a section of this record on a faster time-base; individual m.i.p.s.c.s are marked with short vertical lines.

When the bathing solution was changed to one that contained no calcium and more magnesium to prevent calcium entry through calcium channels, presynaptic depolarization had no effect on transmitter release (Fig. 1c). When nitr-5 was photolysed to raise presynaptic intracellular free calcium concentration, transmitter release increased sharply (Fig. 1e). During neurosecretion, the presynaptic neuron is subjected to another voltage ramp (c). M.i.p.s.c. frequency remained constant throughout this ramp, indicating that the presynaptic potential does not affect transmitter release when presynaptic calcium is held constant.

The results of this experiment are plotted graphically in Fig. 2. The filled circles in Fig. 2a show m.i.p.s.c. frequency to two ramps generated immediately after attaching the patch pipette to the presynaptic neuron in the presence of external calcium. Transmitter release occurred only when the presynaptic neuron was depolarized to $+15$ mV or more. The large variation in m.i.p.s.c. frequency is due to the statistical nature of the release of transmitter quanta at chemical synapses². Depolarization-evoked release was observed only when the bathing medium contained calcium ($N=13$). In three other experiments, we waited after attaching the nitr-5 pipette to cell B5. Transmitter release to depolarization began to drop in five minutes, and was greatly reduced after 10 to 15 minutes, as expected from the buffering action of nitr-5. Thus, we had to work quickly to measure transmitter release to voltage ramps within five minutes of patching cell B5, before most calcium influx was chelated by nitr-5.

Depolarization evoked no transmitter release after changing to a calcium-free medium and allowing nitr-5 to fill the presynaptic cell ($N=4$) (Fig. 2a, dashed line). M.i.p.s.c. frequency remained at virtually zero. The open circles show m.i.p.s.c.

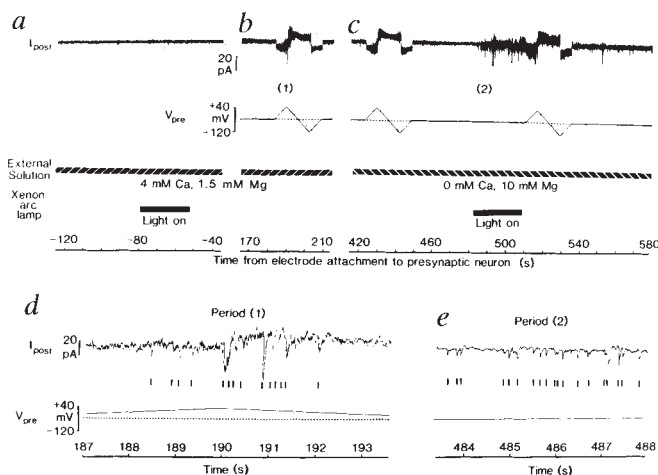


Fig. 1 Selected records from an experiment on a three-day-old B5-B19 neuron pair. The top trace shows postsynaptic (B19) current under voltage clamp and the second trace shows clamped presynaptic (B5) membrane potential. The bottom line shows time from attachment of nitr-5 pipette to neuron B5. Composition of external solution and light exposures are also marked in a to c. a, Spontaneous m.i.p.s.c.s before and during light exposure before introducing nitr-5 into neuron B5. b, Effect of presynaptic membrane potential on transmitter release (m.i.p.s.c. frequency) shortly after attaching patch electrode to cell B19 in calcium-containing medium. The increase in noise level is due to electrical coupling of cells B5 and B19 (ref. 23). The noise in presynaptic voltage imposed by the voltage clamp is passed through to the postsynaptic cell, whose clamp current is recorded with AC coupling at high gain. The command voltage generating presynaptic voltage ramps is composed of a sequence of minute steps; the electrical coupling causes these to appear as additional noise in the postsynaptic current record. The jumps in postsynaptic current at inflexions in the presynaptic voltage ramps are also due to this electrical coupling. c, Effects of presynaptic membrane potential on transmitter release in calcium-free, high-magnesium medium before and after raising presynaptic calcium from $0.24 \mu\text{M}$ to $10 \mu\text{M}$ by photolysing nitr-5. The effect of raising presynaptic calcium on transmitter release is also shown. d, e, High-speed excerpts from the records of b and c at the times marked (1) and (2) respectively. Vertical lines mark the occurrence of m.i.p.s.c.s. These were unambiguously identified by their distinctive waveforms when the records were played back at even faster speeds (not shown).

Methods. Cells B5 and B19 were micro-dissected from buccal ganglia of *Helisoma trivolvis* and were plated together in freshly-collected snail haemolymph in 50% Leibowitz-15 salt-free culture medium (Grand Island Biological) plus 40 mM NaCl, 1.7 mM KCl, 4.1 mM CaCl_2 , 1.5 mM MgCl_2 , 10 mM K^+ -HEPES (Sigma) pH 7.3 and $50 \mu\text{g ml}^{-1}$ gentamicin sulphate (Sigma)²³. Snail haemolymph plasma prevented cell adhesion to the plastic petri dish and discouraged neurite outgrowth. After three to seven days, the cell pair was transferred to a polylysine-coated petri dish for 10–30 min to allow firm attachment to the substrate. Postsynaptic current was recorded from B19 using a Dagan Corporation 8900 patch clamp circuit and patch pipettes with 2–3 μm openings and containing 50 mM KCl, 5 mM MgCl_2 , 5 mM EGTA and 5 mM K^+ -HEPES pH 7.3. Presynaptic cell B5 was voltage-clamped with a similar circuit. Its pipette contained 12 mM nitr-5 loaded with 9 mM Ca^{2+} and dissolved in 35 mM KCl, 5 mM MgCl_2 and 5 mM K^+ -HEPES pH 7.3. Nitr-5 was photolysed with an ILC Technology 150 W Cermex xenon arc lamp fitted with a liquid filter to eliminate infrared and far ultraviolet radiation. The intensity of this light at 360 nm was calibrated by determining its photolytic efficiency on dilute solutions of nitr-5 (ref. 21).

frequency after presynaptic intracellular calcium was elevated from its pre-photolysis level of 240 nM to $10 \mu\text{M}$ (calculated as described below) by the light exposure. Transmitter release was now constant over the range of -120 to $+40$ mV, as shown by a least-squares regression fit to this data (Fig. 2a, dotted line). This contrasts sharply with the strong voltage-dependence of transmitter release caused by calcium influx through voltage-activated calcium channels (Fig. 2a). Similar results were obtained in all four cell pairs examined.

One explanation for this result is that after nitr-5 photolysis the synapse is releasing transmitter at its maximum rate and is unable to respond to presynaptic depolarization; another is that some cytoplasmic factor needed to confer voltage-dependence upon transmitter release is washed out of the presynaptic neuron through the patch pipette. These possibilities were tested in four

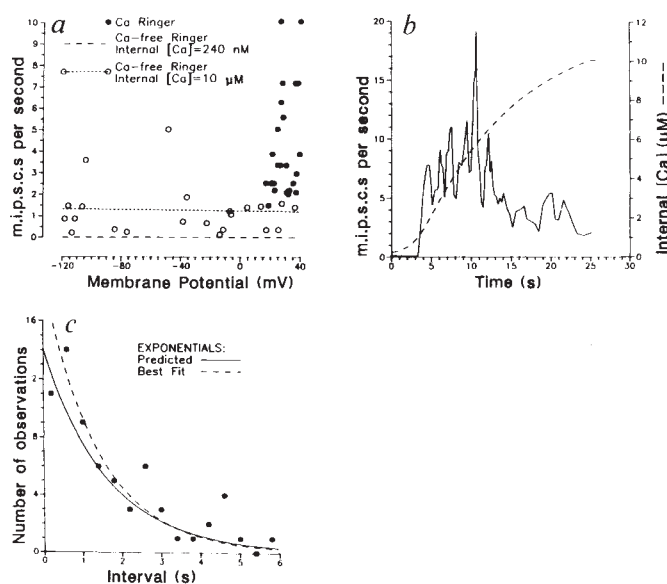


Fig. 2 *a*, Frequency of m.i.p.s.c.s against membrane potential during presynaptic voltage ramps in calcium-containing medium, calcium-free, high-magnesium, medium before nitr-5 photolysis and calcium-free medium after nitr-5 photolysis. Data combined from two ramps under each experimental condition (Fig. 1*b* and *c* shows one of each of these ramps). Intracellular calcium concentration was calculated as described in text. Dashed line represents zero frequency because only one m.i.p.s.c. occurred during ramps in calcium-free medium before nitr-5 photolysis. The dotted line is a regression line through data points after nitr-5 photolysis. *b*, Observed m.i.p.s.c. frequency and calculated presynaptic free calcium concentration during 25-second light exposure. M.i.p.s.c. frequency was smoothed using the moving bin method³¹ to average successive groups of four intervals. Because this method blurs the initial sharp rise in m.i.p.s.c. frequency, it was not applied to the first three points. *c*, Histogram of m.i.p.s.c. intervals after nitr-5 photolysis raised presynaptic calcium to 10 μ M and evoked transmitter release. Solid line, expected distribution of m.i.p.s.c. intervals assuming a Poisson release process (see text). Dotted line, least-squares exponential fitted to data points.

control experiments in which calcium remained present in the bathing saline. Presynaptic depolarization, which admitted additional calcium into the cell, caused an additional increase in transmitter release after presynaptic liberation of caged calcium by photolysis of nitr-5 to its low-affinity form. Only when calcium influx was prevented did depolarization fail to accelerate transmitter release. Taken together, our results indicate that presynaptic voltage does not directly evoke the release of neurotransmitter.

The m.i.p.s.c. frequency during the 25-second exposure to the light was measured (Fig. 2*b*). The effect of light on intracellular calcium levels was calculated for a synapse located midway between the front surface of the B5-B19 cell pair facing the light source and the back surface of the cell pair resting on the substrate²⁸. Briefly, this calculation involves solution of the diffusion equation for calcium and low- and high-affinity forms, both free and complexed, of nitr-5 and native buffer, as well as photolysis of nitr-5 by light as a function of distance below the front surface and simultaneous equilibration of all calcium buffers at each point in presynaptic cytoplasm. This calculation can be done for a calibrated light source knowing the initial concentrations, light absorbances, calcium affinities, and diffusion constants of all diffusing species.

These calculations indicate that transmitter release increases when free calcium concentration rises from 0.25 μ M to about 1.5 μ M (observed in nine of 12 experiments). Maximum m.i.p.s.c. frequency is reached in five to 10 seconds after calcium

rises to about 6 μ M. The rate at which this level of intracellular calcium releases transmitter is almost as high as that caused by depolarizations of the same magnitude as action potentials in normal calcium-containing medium (P.G.H., unpublished observations). A further increase in presynaptic free calcium concentration to about 10 μ M is accompanied by a decrease in m.i.p.s.c. frequency. This may be due to depletion of transmitter stores, as this synapse is prone to synaptic depression (P.G.H., unpublished observations). Alternatively, high levels of presynaptic calcium could have a secondary blocking effect on transmitter release²⁹, or the nitrosobenzophenone form of nitr-5 might have a mild inhibitory effect on secretion.

After exposure to light, transmitter release persists for several minutes. These m.i.p.s.c.s were used to obtain a histogram of intervals between quanta released. The results fit an exponential distribution, as expected for a Poisson random process³⁰ (Fig. 2*c*). The best-fit exponential (dashed line) is close to the expected exponential, whose time constant is the mean m.i.p.s.c. interval, and whose area is that of the histogram (solid line).

These experiments lead us to draw several conclusions. 1) Transmitter release can be evoked by raising intracellular calcium above 1 μ M. 2) When presynaptic free calcium is raised to about 5–10 μ M, the rate of transmitter release is nearly that evoked by depolarization to the same extent as action potentials. This calcium level is nearly as high as that estimated to occur at release sites near calcium channels during action potentials⁵. 3) No transmitter release occurs when the presynaptic membrane is depolarized in a calcium-free medium and presynaptic calcium concentration is low. 4) Presynaptic membrane potentials between -120 and +40 mV have no effect on the transmitter release evoked by elevating presynaptic calcium in the absence of calcium influx. 5) Action potentials evoke transmitter release only by opening voltage-dependent calcium channels to allow calcium in and raise its concentration at transmitter release sites. This contrasts sharply with the hypothesis that action potentials directly activate transmitter release in addition to admitting calcium to release sites.

We are grateful to Drs Roger Tsien and Stephen Adams for providing a sample of nitr-5 and for discussion. We thank Liz Sievers for technical assistance and Dr Luca Landò for participating in preliminary experiments. This work was supported by the National Institutes of Health.

Received 21 June; accepted 12 August 1988.

1. Llinás, R., Steinberg, I. Z. & Walton, K. *Biophys. J.* **33**, 323–352 (1981).
2. Katz, B. *The Release of Neural Transmitter Substances* (ed. Thomas, C. C., Springfield, Ill., 1969).
3. Miledi, R. *Proc. R. Soc. Lond. B* **183**, 421–425 (1973).
4. Simon, S. M. & Llinás, R. *Biophys. J.* **48**, 485–498 (1985).
5. Fogelson, A. L. & Zucker, R. S. *Biophys. J.* **48**, 1003–1017 (1985).
6. Dudel, J. *Neurosci. Lett.* **41**, 133–138 (1983).
7. Dudel, J., Parnas, I. & Parnas, H. *Pflügers Arch. ges. Physiol.* **399**, 1–10 (1983).
8. Dudel, J. *Pflügers Arch. ges. Physiol.* **402**, 225–234 (1984).
9. Dudel, J. *Pflügers Arch. ges. Physiol.* **402**, 235–243 (1984).
10. Parnas, H. & Segel, L. A. *J. theor. Biol.* **107**, 345–365 (1984).
11. Parnas, I., Dudel, J. & Parnas, H. *Neurosci. Lett.* **50**, 157–162 (1984).
12. Dudel, J. *Pflügers Arch. ges. Physiol.* **406**, 449–457 (1986).
13. Parnas, I., Parnas, H. & Dudel, J. *Pflügers Arch. ges. Physiol.* **406**, 121–130 (1986).
14. Parnas, I., Parnas, H. & Dudel, J. *Pflügers Arch. ges. Physiol.* **406**, 131–137 (1986).
15. Parnas, I. & Parnas, H. *J. Physiol., Paris* **81**, 289–305 (1986).
16. Zucker, R. S. & Landò, L. *Science* **231**, 574–579 (1986).
17. Zucker, R. S., Landò, L. & Fogelson, A. L. *J. Physiol., Paris* **81**, 237–245 (1986).
18. Zucker, R. S. *Biophys. J.* **51**, 347–350 (1987).
19. Zucker, R. S. & Fogelson, A. L. *Proc. natn. Acad. Sci. U.S.A.* **83**, 3032–3036 (1986).
20. Augustine, G. J., Charlton, M. P. & Smith, S. J. *J. Physiol., Lond.* **367**, 163–181 (1985).
21. Tsien, R. & Zucker, R. S. *Biophys. J.* **50**, 843–853 (1986).
22. Adams, S. R., Kao, J. P. Y., Grynkiewicz, G., Minta, A. & Tsien, R. Y. *J. Am. chem. Soc.* **110**, 3212–3220 (1988).
23. Haydon, P. G. *J. Neurosci.* **8**, 1032–1038 (1988).
24. Bennett, M. R. & Pettigrew, A. G. *J. Physiol., Lond.* **241**, 547–573 (1974).
25. Dennis, M. J. *J. Physiol., Lond.* **244**, 683–702 (1975).
26. Xie, Z.-p. & Poo, M.-m. *Proc. natn. Acad. Sci. U.S.A.* **83**, 7069–7073 (1986).
27. Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. *Pflügers Arch. ges. Physiol.* **391**, 85–100 (1981).
28. Landò, L. & Zucker, R. S. *J. gen. Physiol.* (in the press).
29. Adams, D. J., Takeda, K. & Umbach, J. A. *J. Physiol., Lond.* **369**, 145–159 (1985).
30. Fatt, P. & Kaatz, B. *J. Physiol., Lond.* **117**, 109–128 (1952).
31. Rahamimoff, R. & Yaari, Y. *J. Physiol., Lond.* **228**, 241–257 (1973).

Substitution of murine for human CD4 residues identifies amino acids critical for HIV-gp120 binding

Linda K. Clayton^{*†}, Rebecca E. Hussey^{*},
Ruth Steinbrich^{*}, Hema Ramachandran^{*},
Yasmin Husain^{*} & Ellis L. Reinherz^{*‡}

^{*} Laboratory of Immunobiology and Departments of [†]Pathology and [‡]Medicine, Harvard Medical School, 44 Binney Street, Boston, Massachusetts 02115, USA

Human CD4 is the receptor for the gp120 envelope glycoprotein of human immunodeficiency virus and is essential for virus entry into the host cell¹⁻⁸. Sequence analysis of CD4 has suggested an evolutionary origin from a structure with four immunoglobulin-related domains^{9,10}. Only the two NH₂-terminal domains are required to mediate gp120 binding¹¹⁻¹³. The extracellular segment of murine CD4 has an overall 50% identity with its human counterpart¹⁴ at the amino-acid level, but fails to bind gp120¹⁵. To define those residues of human CD4 critical for gp120 binding, we have taken advantage of this species difference and substituted all non-conserved murine for human CD4 residues between amino-acid positions 27-167. We used oligonucleotide-directed mutagenesis to create each of 16 individual mutant human CD4 molecules containing from 1-4 amino-acid substitutions. Introduction of as few as three amino acids into corresponding positions of human CD4 abrogates gp120 binding. Furthermore, these critical residues are located in domain I with a contribution from domain II. Modelling studies using the three-dimensional coordinates of the V κ Bence-Jones REI homodimer localize the site in domain I to the C' β strand within CDR2 but projecting away from the homologues of principle antigen-binding regions CDR 1 and 3.

To characterize the residues of the CD4 structure involved in gp120 binding, we used a COS-1 cell expression system and the previously described¹⁶ anchor-minus CD4 construct termed T_{4_{ex1}}. As shown in Fig. 1a, the 370-amino-acid T_{4_{ex1}} protein comprises three intrachain disulphide bonded domains (a domain is defined as residues between and including 20 amino-acid residues to either side of the cysteines) and one domain (III) which lacks cysteine residues but is immunoglobulin-like in the same way as its counterparts⁹. Nanomolar concentrations of T_{4_{ex1}} inhibit gp120-transmembrane CD4 interaction, syncytium formation and human immunodeficiency virus (HIV) infection by binding to gp120-expressing cells¹⁶. The T_{4_{ex1}} construct was subcloned into the vector CDM8 (ref. 17) and transfected into COS-1 cells. Supernatants from metabolically labelled transfected cells were tested by immunoprecipitation with an anti-CD4 monoclonal antibody (19Thy5D7) and the resulting precipitate subjected to SDS-PAGE. Figure 1b, lane 3 shows the presence of a CD4-derived molecule of relative molecular mass 50,000 (50K) in transfected COS-1 cell supernatants. The same molecule is coprecipitated from COS-1 supernatants with an anti-gp120 monoclonal antibody after preincubation of the supernatant with gp120 (Fig. 1b, lane 5). The specificities of these reactions are demonstrated by the facts that an irrelevant antibody (anti-T8) fails to precipitate T_{4_{ex1}} (Fig. 1b, lane 1) and that no CD4 band is detected with anti-gp120 antibody in the absence of gp120 (Fig. 1b, lane 6).

Earlier studies using either CD4 DNA truncation^{11,12} or proteolytic digestion¹³ demonstrated that the residues critical for gp120 interaction reside in domains I and/or II exclusively. Similarly, the COS-1 cell-derived product of a T_{4_{ex1}} protein truncated after amino-acid residue 182 (by insertion of a stop codon in the complementary DNA sequence) is precipitated as a 20K protein by anti-CD4 antibody and binds to gp120 (Fig. 1b, lanes 4 and 7 respectively). We therefore further analysed

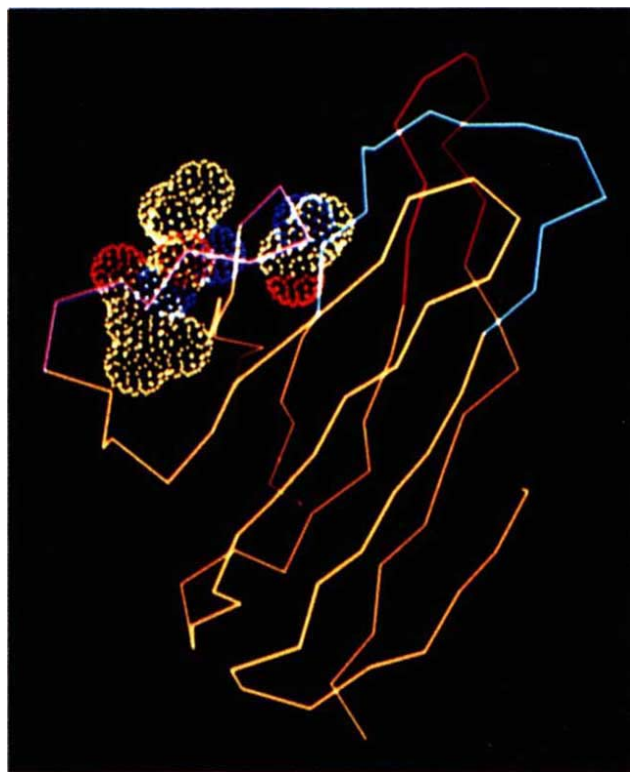


Fig. 3 The location of residues in domain I of CD4 predicted to be involved in gp120 binding. Alpha-carbon skeleton of a V κ domain derived from coordinates of the Bence-Jones homodimer REI (ref. 30). One member of the dimer is shown. Spheroids indicate residues in REI corresponding to the region of amino-acid substitutions in mutant M3. The alignment of CD4 and REI is similar to that of Maddon *et al.*¹⁹ beginning at residue 8 of REI. Diagrams were generated using an Evans and Sutherland PS-300 and the program Frodo³¹. The dyad of the immunoglobulin variable domain is vertical. CDR1, 2 and 3 are delineated in the colours blue, purple and red respectively. The NH₂-terminus is readily visible in this view and the COOH-terminus is at the bottom.

the CD4-gp120 interaction by creating 35 amino-acid substitutions encompassing all non-conservative mouse-human species differences within the first two domains of CD4 between residues 27 and 167. The NH₂-terminal CD4 amino acids were not considered here because an NH₂-terminal synthetic peptide failed to block gp120 binding even at millimolar concentrations (our unpublished data). For each substitution we replaced an amino acid of the human sequence with that amino acid found in the equivalent position of the murine CD4 sequence¹⁴. As the latter does not bind gp120, we anticipated that some murine substitutions would abrogate human CD4-gp120 interaction. As shown in Table 1, 15 oligonucleotides were used in a standard site-directed mutagenesis protocol (see Fig. 1 legend) to produce 16 different versions of the human CD4 molecule containing from 1-4 substitutions each. The positions of these substitutions are listed in Table 1 and diagrammatically mapped in Fig. 1a. All 16 CD4 mutants were assayed after transfection into COS-1 cells by immunoprecipitation with anti-CD4 monoclonal antibody and by gp120 co-precipitation with anti-gp120. Immunoprecipitation of the original T_{4_{ex1}} and four representative mutants is shown in Fig. 2a. In addition to T_{4_{ex1}}, each of the mutants M5, M10, M7 and M3 react with the anti-CD4 monoclonal antibody 19Thy5D7. As shown in Table 1, 15 of the 16 mutants react with anti-CD4 antibody. Only mutant M14 fails to react and is furthermore unreactive with OKT4A, a second monoclonal antibody directed at a different CD4 epitope.

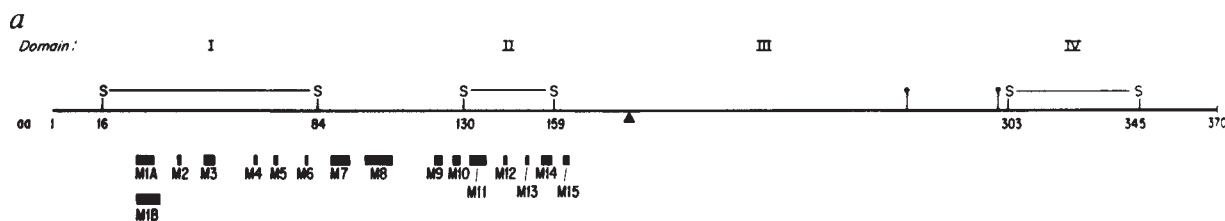
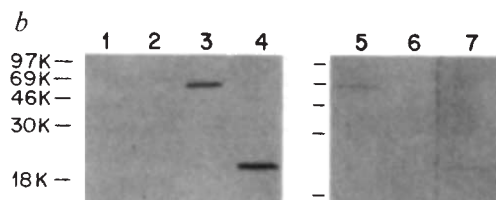


Fig. 1 CD4-derived proteins which bind to gp120. *a*, A schematic diagram of CD4 protein T4_{ex1} showing the four immunoglobulin-like domains, three disulphide bonds and two potential glycosylation sites⁹. Numbering of amino acids is according to ref. 16. The positions of 16 mutations (further delineated in Table 1) are shown below the line. The triangle indicates a stop codon introduced by site-directed mutagenesis to create a protein containing only the first 182 amino acids. *b*, Anti-CD4 immunoprecipitation and anti-gp120 co-precipitation of T4_{ex1} and a truncated 182-amino-acid version of CD4 from supernatants of COS-1 cells labelled with [³⁵S] cysteine and transfected with the CD4 constructs. Lane 1,



immunoprecipitation of supernatant from COS-1 cells transfected with the T4_{ex1}-containing plasmid and immunoprecipitated with anti-T8 (21Thy2D3) (control); lane 2, immunoprecipitation of supernatant from COS-1 cells transfected with the 182-amino-acid truncation using the control anti-T8 antibody; lane 3, immunoprecipitation of T4_{ex1} with anti-CD4 antibody (19Thy5D7); lane 4, immunoprecipitation of the 182-amino-acid truncation of T4_{ex1} with anti-CD4; lane 5, co-precipitation of T4_{ex1} with anti-gp120 (Dupont, see ref. 13) in the presence of gp120; lane 6, co-precipitation of T4_{ex1} with anti-gp120 in the absence of gp120; lane 7, co-precipitation of the 182-amino-acid truncation of T4_{ex1} with anti-gp120 in the presence of gp120. All gel samples are non-reduced. The molecular weight markers are phosphorylase B (97.4K), bovine serum albumin (69K), ovalbumin (46K), carbonic anhydrase (30K) and lactoglobulin A (18.4K).

Methods. The 1.17-kilobase(kb) T4_{ex1} fragment was excised using *Bam*HI from pAC 373/T4_{ex1} (ref. 16), blunted using the Klenow fragment of DNA polymerase I, ligated to *Xho*I linkers (New England Biolabs) and subcloned in the *Xho*I site of the vector CDM8 (ref. 17). For transfection of CDM8 constructs into COS cells, 2–3 × 10⁶ cells are plated in 100 × 15 cm dishes in RPMI 1640 (Gibco) containing 10% fetal calf serum (FCS). Twelve to 24 h later, 45 µg plasmid DNA are added to 2.5 ml RPMI and mixed with 2.5 ml RPMI-800 µg ml⁻¹ DEAE dextran, then added to the washed COS cells. After ~2 h at 37 °C, the cells are washed and then incubated in RPMI containing 2% FCS, 1% glutamine, 1% penicillin-streptomycin, 10 µg ml⁻¹ gentamycin and 150 µM chloroquine for 3 h. The cells are incubated at 37 °C for 2 d in RPMI 10% FCS. For metabolic labelling, the transfected COS cells (2 d after transfection) are incubated for 1 h in 5 ml RPMI minus cysteine containing 10% FCS. After removal of media, the cells are incubated in RPMI minus cysteine containing 10% dialysed FCS and 100 µCi ml⁻¹ of [³⁵S]cysteine for 5–6 h at 37 °C. The supernatants are removed, centrifuged at 200g for 10 min and dialysed versus PBS/0.025% azide/10 mM cold cysteine overnight at 4 °C. For immunoprecipitation, 5 ml of the dialysed [³⁵S]cysteine-labelled supernatant is precleared by a 45 min incubation at 4 °C with 20 µl anti-T8 antibody (21Thy2D3) on Affigel-10 (Biorad) beads (~5 mg antibody per ml beads). The precleared supernatant is then incubated with 20 µl anti-CD4 (19Thy5D7) on Affigel-10 beads for 3 h at 4 °C. The beads are washed once in 10 ml 10 mM Tris, pH 6.8/0.1% Triton X-100/0.1% SDS/0.5% deoxycholate, once in 1 ml of the same buffer and once in 1 ml 0.1 M glycine, pH 5/0.1% Triton X-100 and then eluted with 35 µl 0.1 M glycine, pH 2/0.1% Triton X-100 and neutralized with 6 µl 1 M Tris, pH 7.6. The sample is run on a 0.75 or 1.5 mm 12.5% mini-polyacrylamide-SDS gel under non-reducing conditions. The gel is fixed, dried and autoradiographed at ~70 °C from 1–7 days. Immunoprecipitation with anti-CD8 was as above, except that 20 µl anti-CD8 on Affigel-10 beads is used. For co-precipitation with gp120 (gift from Drs Matthews and Bolognesi, see ref. 13), 0.5 ml of labelled supernatant is incubated with 67 ng native gp120 for 2 h at 37 °C. Five hundred ng anti-gp120 (Dupont) and 10 µl rabbit anti-mouse IgG Sepharose 4B beads are added and rotated for 2 h at 4 °C. The beads are washed once in 10 ml and once in 1 ml cold phosphate buffered saline, eluted and the sample run in SDS-PAGE as above.

The CD4 protein (182 amino acids long) was created using the thionucleotide method of oligonucleotide site-directed mutagenesis^{27–29}. The *Xho*I insert of T4_{ex1} was excised from CDM8, blunted with the Klenow fragment of DNA polymerase I, ligated to *Xba*I linkers (New England Biolabs) and subcloned into M13mp18. Single-stranded DNA was prepared as a template and mutagenesis was carried out according to the manufacturer's recommendations (Amersham). For the 182-amino-acid truncation, the oligonucleotide 5' GAAGGCCTAAAGCATAG 3' was synthesized using standard cyanoethyl phosphoramidite chemistry. The termination codon which converts the serine (TCC) at amino acid 183 to a stop codon is underlined. The presence of the mutation was confirmed by sequencing the M13mp18-T4 construct and mini-preps of the replicative form of the mutation-containing DNA were carried out. The mutated insert was excised with *Xba*I and ligated into the *Xba*I site of CDM8. The presence of the mutation was then directly confirmed by sequencing the CDM8-T4 insert using the double-stranded DNA as a template.

Thirteen of the 16 mutants bound gp120 in a manner equivalent to T4_{ex1} as judged by the co-precipitation assay. Figure 2b demonstrates that T4_{ex1}, M5, M10 and M7 are all co-precipitated by anti-gp120 in the presence of gp120. Overall, we observed a 2–3-fold experimental variation in co-precipitation with gp120 (T4_{ex1} versus M5 in Fig. 2b); among gp120-binding CD4 proteins, however, a positive signal was observed in every experiment (using a minimum of 2–3 separate transfections). In contrast, although M3 is recognized by anti-CD4 antibody, it fails to bind to gp120 (Fig. 2b). In addition, M9 (Table 1) has a substantially reduced gp120-binding capacity, although anti-CD4 monoclonal antibody immunoprecipitates a band of identical size and intensity to T4_{ex1}. M3 contains three amino-acid substitutions in human CD4 domain I at positions 48, 50 and 51. One or more of these changes clearly abrogates the ability of CD4 to bind to gp120. M9 contains three amino-acid substitutions in domain II of CD4 at position 121–123.

Thus, alteration of a few residues in either CD4 domain I or domain II can abrogate or reduce gp120 binding respectively.

In addition, M14 demonstrates reduced binding to gp120 (Table 1). Given that it fails to bind to the two anti-CD4 monoclonal antibodies examined, we cannot rule out the possibility that the three substitutions in M14 at positions 155, 156 and 158 somehow decrease the expression of this mutant CD4 protein. Nevertheless, it is more likely that these substitutions have destroyed both the gp120 binding site and the epitopes recognized by the two monoclonal antibodies, perhaps through a general disruption of the CD4 protein's three-dimensional structure because translation of RNA transcribed *in vitro* from M14 gave results identical to T4_{ex1}-transcribed RNA (data not shown).

The contribution of CD4 domain I to gp120 binding was recognized previously in studies of the T4_{ex1} polypeptide produced in a baculovirus system in conjunction with proteolytic

Table 1 Production and analysis of CD4 site-directed mutants

Mutant	Oligonucleotide used for mutagenesis	Amino acid change	Anti-CD4 immunoprecipitation	Anti-gp120 co-precipitation
M1A	Mouse substitution 223 CAA-TTC-ACC-TGG-AAA-TTC-TCC-GAC-CAG-AGA-AAG 255 Human <i>T</i> _H <i>F</i> _N <i>D</i> _N <i>I</i> [*]	27 H to T 30 N to F 32 N to D	+	+
M1B	223 CAA-TTC-ACC-TGG-AAA-TTC-TCC-GAC-CAG-AGA-AAG 255 <i>T</i> _H <i>F</i> _N <i>D</i> _N <i>R</i> _I	27 H to T 30 N to F 32 N to D 34 I to R	+	+
M2	261 G-GGA-AAT-CAC-GGC-TCC 276 <i>H</i> _Q	40 Q to H	+	+
M3	283 ACT-AAA-GGT-GGA-TCC-CCG-AGT-AAT-GAT-CG 311 <i>G</i> _P <i>P</i> _K <i>S</i> _L	48 P to G 50 K to P 51 L to S	+	-
M4	335 GG-GAC-AAA-GGA-AAC-TTC 351 <i>K</i> _Q	64 Q to K	+	+
M5	355 CTG-ATC-ATC-AAT-AAAG-CTT-AAG 375 <i>N</i> _K <i>K</i> _N	72 K to N 73 N to K	+	+
M6	382 GAC-TCA-CAG-ACT-TAC-ATC 399 <i>Q</i> _D	80 D to Q	+	+
M7	406 GTG-GAG-AAC-CCG-AAG-GAG-GAG-GTG-GAA-TTG-C 436 <i>N</i> _D <i>R</i> _Q <i>E</i> _Q	88 D to N 89 Q to R 94 Q to E	+	+
M8	436 CTA-GTG-TTC-AAA-TTG-ACT-GCC-AAC-CCT-GAC-ACC-AGC-CTG-CTT-C 478 <i>K</i> _G <i>P</i> _S <i>S</i> _H	99 G to K 104 S to P 107 H to S	+	+
M9	499 ACC-TTG-GAG-AGC-AGC-AAG-GTT-AGT-AGC-CCC 528 <i>S</i> _P <i>K</i> _P <i>V</i> _G	121 P to S 122 P to K 123 G to V	+	±‡
M10	520 AGT-AGC-CCC-CTA-ACG-GAA-TGT-AGG 543 <i>L</i> _S <i>T</i> _V <i>E</i> _Q	127 S to L 128 V to T 129 Q to E	+	+
M11	534 G-CAA-TGT-AGG-CAT-AAA-AGG-GGT-AAA-GTC-ATA-CAG-GG 569 <i>H</i> _S <i>K</i> _P <i>V</i> _N	132 S to H 133 P to K 137 N to V	+	+
M12	570 G-GGG-AAG-GTC-CTC-TCC-G 586 <i>V</i> _T	143 T to V	+	+
M13	590 CT-CAG-CTG-CCG-CTC-CAG-G 607 <i>R</i> _E	150 E to R	+	+
M14	606 G-GAT-AGT-GAC-TTC-TGG-AAT-TGC-ACT-GTC 633 <i>D</i> _G <i>F</i> _T <i>N</i> _T	155 G to D 156 T to F 158 T to N	-†	±‡
M15	626 GC-ACT-GTC-ACG-CTG-GAC-CAG-AAG 648 <i>T</i> _L <i>L</i> _Q <i>D</i> _N	162 L to T 163 Q to L 164 N to D	+	+

* Two mutants were recovered from the mutagenesis using this oligonucleotide; one contained mutations at amino acid 27, 30 and 32 but not 34, and the second contained all four changes. These two mutants were transfected separately. Letters in italics above and below nucleotide sequencer refer to amino acids in single-letter code.

† M14 was also negative when tested for immunoprecipitation with anti-CD4 monoclonal OKT4A.

‡ A very faint 50 K band (~10-fold less intense than T_{4ext}) was observed on co-precipitation with gp120. Mutagenesis, immunoprecipitation and co-precipitation procedures are described in the legend to Fig. 1.

fragmentation analysis, microsequencing and a specific CD4-gp120 binding assay¹³. There it was shown that disruption of the peptide bond at lysine 72 by tryptic cleavage destroyed CD4-gp120 interaction without inducing any detectable alterations in other domains of CD4. Furthermore, reduction of intrachain disulphide bonds in the CD4 molecule also abrogated high-affinity gp120 binding, thereby strongly implying that the binding site for gp120 is complex and depends on the stabilized CD4 structure. Whether the introduced domain I and II mutations affect gp120 contact residues themselves or the tertiary structure around the contact residues is not clear at present. Nevertheless, the ability of a synthetic peptide comprising amino acids 23–56 to inhibit syncytium formation at 10⁻⁴ M (ref. 18) may support the notion that residues 48, 50 and/or 51 contribute to the gp120 binding site. Undoubtedly the reduction in affinity of the synthetic peptide relative to CD4 by five orders of magnitude reflects a substantial requirement for tertiary structure in gp120 binding.

Given the similarities of CD4 to immunoglobulin¹⁹, we have modelled CD4 domain I based on the known three-dimensional coordinates of the V κ Bence-Jones homodimer, REI, and determined the relative positions of the M3 mutations at residues 48, 50 and 51 of CD4. We previously predicted that each of three tryptic cleavage sites in domain I were exposed at the surface¹³, thus supporting the validity of the immunoglobulin model for CD4. Figure 3 shows the α -carbon skeleton of one chain of the REI homodimer. The region of residues in REI corresponding to the mutated CD4 residues that abrogate gp120 binding are highlighted by spheroids. This region corresponds to the C' β -strand unique to V domains that connects the two β -sheets and comprises part of CDR2 (residues 50–56)²⁰. Because the alignment between REI and CD4 requires a gap in this segment, we do not wish to imply that the CD4 α -carbon skeleton follows an identical course in this region. Nevertheless, it is very likely that the CD4 sequence will also loop out and be exposed to solvent. Furthermore, note that this site is distinct from the

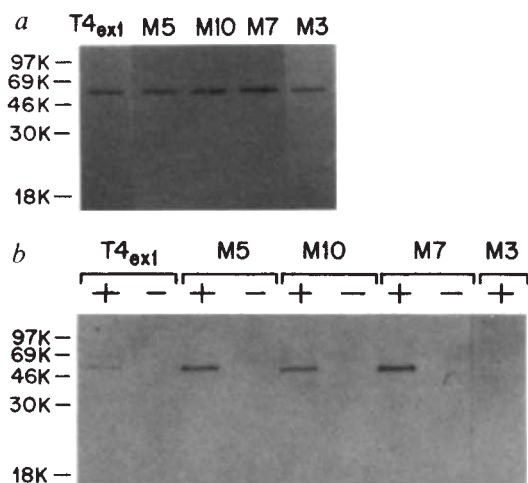


Fig. 2 *a*, Anti-CD4 immunoprecipitation of [35 S] cysteine labelled supernatants from COS-1 cells transfected with T4_{ex1}, M5, M10, M7 and M3. *b*, Anti-gp120 co-precipitation of [35 S] cysteine labelled supernatants from COS-1 cells transfected with T4_{ex1}, M5, M10, M7 and M3. Precipitations were carried out in the presence (+) or absence (-) of gp120.

Methods. Construction of mutants, transfections, immunoprecipitations and co-precipitations were carried out as described in the Fig. 1 legend. The presence of each mutation was confirmed by directly sequencing the plasmid DNA used for individual transfections.

segments equivalent to CDR1 (residues 23–34) and CDR3 (residues 89–97), the main antigen contact regions²¹ which are located at the top (right) of Fig. 3. If CDR1 and three homologous regions form a binding site for class II major histocompatibility complex products, the putative natural ligands of CD4^{22–26}, one can speculate that gp120 may be incapable of inhibiting class II recognition events. Based on the above analysis, one further prediction would be that if the gp120 binding site does involve residues in the region analogous to the C' strand of REI, it might also include or be conformationally altered by residues in CD4 domain II adjacent to this region. Perhaps the M9 and/or M14 mutations are localized to such sites.

We thank Professor Stephen Harrison for advice and assistance with computer modelling, Drs Peter Sayre and George Tarr for discussion, Monica Sieh and Dr Amita Vaid for technical assistance. This work was supported in part by NIH.

Note added in proof: Since we submitted this paper, Peterson and Seed³² have published a mutational analysis of HIV binding sites on human CD4 in which mutations in the same region as those in M3 affected virus binding and syncytium formation. In addition, Landau *et al.*³³ have used chimaeric mouse-human CD4 molecules to study HIV gp120 binding and found that the minimum region of human CD4 necessary for gp120 binding includes the region identified in M3. Full gp120 binding required additional human amino acids including those identified in M9 and M14.

Received 28 June; accepted 5 August 1988.

- Klatzmann, D. *et al.* *Science* **225**, 59–63 (1984).
- Dalgleish, A. G. *et al.* *Nature* **312**, 763–766 (1984).
- Sattentau, Q., Dalgleish, A., Weiss, R. & Beverley, P. C. L. *Science* **234**, 1120–1123 (1986).
- McDougal, J. S. *et al.* *J. Immun.* **137**, 2937–2944 (1986).
- McDougal, J. S. *et al.* *Science* **231**, 382–385 (1986).
- Maddon, P. J. *et al.* *Cell* **47**, 333–348 (1986).
- Sodroski, J., Goh, W. C., Rosen, C., Campbell, K. & Haseltine, W. A. *Nature* **322**, 470–474 (1986).
- Lifson, J. *et al.* *Nature* **323**, 725–728 (1986).
- Clark, S., Jeffries, W. A., Barclay, A. N., Gagnon, J. & Williams, A. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1649–1653 (1987).
- Littman, D. R. & Gettner, S. N. *Nature* **325**, 453–455 (1987).
- Trautwein, A., Wolfgang, L. & Karjalainen, K. *Nature* **331**, 84–86 (1988).
- Berger, E. A., Fuerst, T. R. & Moss, B. *Proc. natn. Acad. Sci. U.S.A.* **85**, 2357–2361 (1988).
- Richardson, N. E. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **85**, 6102–6106 (1988).
- Maddon, P. J. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **84**, 9155–9159 (1987).
- McClure, M. O. *et al.* *Nature* **330**, 487–489 (1987).

- Hussey, R. E. *et al.* *Nature* **331**, 78–81 (1988).
- Seed, B. & Aruffo, A. *Proc. natn. Acad. Sci. U.S.A.* **84**, 3365–3369 (1987).
- Jameson, B. A. *et al.* *Science* **240**, 1335–1339 (1988).
- Maddon, P. *et al.* *Cell* **42**, 93–104 (1985).
- Williams, A. F. & Barclay, A. N. *Rev. Immun.* **6**, 381–405 (1988).
- Amit, A. G., Mariuzza, R. A., Phillips, S. E. V. & Poljack, R. J. *Nature* **313**, 156–158 (1985).
- Krensky, A. M., Reiss, C. S., Mier, J. W., Strominger, J. L. & Burakoff, S. J. *Proc. natn. Acad. Sci. U.S.A.* **79**, 2365–2369 (1982).
- Meuer, S. C., Schlossman, S. F. & Reinherz, E. L. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4395–4399 (1982).
- Biddison, W., Rao, P., Talle, M. A., Goldstein, G. & Shaw, S. J. *J. exp. Med.* **156**, 1065–1076 (1982).
- Marrack, P. *et al.* *J. exp. Med.* **158**, 1077–1091 (1983).
- Doyle, C. & Strominger, J. L. *Nature* **330**, 256–259 (1987).
- Taylor, J. W. *et al.* *Nucleic Acids Res.* **13**, 8749–8764 (1985).
- Taylor, J. W., Ott, J. & Eckstein, F. *Nucleic Acids Res.* **13**, 8765–8785 (1985).
- Nakayama, K. & Eckstein, F. *Nucleic Acids Res.* **14**, 9679–9698 (1986).
- Epp, O., Lattman, E. E., Schiffer, M., Huber, R. & Palm, W. *Biochemistry* **14**, 4943–4952 (1975).
- Jones, T. A. *J. appl. Crystallogr.* **11**, 268 (1978).
- Peterson, A. & Seed, B. *Cell* **54**, 65–72 (1988).
- Landau, N. R., Warton, M. & Littman, D. R. *Nature* **334**, 159–162 (1988).

Infection of rabbits with human immunodeficiency virus

Gaetano Filice*, Paolo M. Cereda† & Oliviero E. Varnier‡

* Institute of Infectious Diseases, IRCCS "San Matteo" and
† Institute of Microbiology, School of Medicine, 27100 Pavia, Italy
‡ Laboratory of Human Retrovirology, Institute of Microbiology,
School of Medicine, 16132 Genova, Italy

An important requirement for the development of a vaccine against the human immunodeficiency virus (HIV-1)¹, the causative agent of AIDS², is a readily available animal model that would allow possible immunogens to be evaluated³. The only species to have been infected with HIV-1 so far is the chimpanzee^{4–7}. However, the scarcity of this animal and its designation as an endangered species⁸ place severe restrictions on its use as an animal model. Attempts to infect mice, rats, hamsters, guinea-pigs, musk shrews, and rabbits with HIV-1 or infected cells have all been unsuccessful⁹. We now report that the intraperitoneal inoculation of rabbits with HIV-1 or chronically infected H9 cells consistently induces a persistent infection.

Aseptic thioglycollate peritonitis was induced in a group of nine rabbits when they were 35–40 days old. After 4–5 days, three rabbits were injected intraperitoneally (i.p.) in the presence of interleukin-2 (IL-2) with 1×10^7 H9 cells infected with the HTLV-III_{B10} (ref. 10) strain of HIV-1. Three rabbits were inoculated with 5 ml of cell-free supernatant of H9/HTLV-III_B cells (1×10^5 infectious units per ml). Two control animals received uninfected H9 cells and another complete growth medium only (Table 1).

Two weeks after the inoculation, HIV-1 antibodies were detectable by an enzyme-linked immunosorbent assay (ELISA) in sera from all the rabbits treated with material containing HIV-1. HIV-1 antigens were also detected in serum samples. All the HIV-1-treated rabbits were seropositive one year after the inoculation (Table 1). Immunoglobulin G (IgG) antibody titres were relatively high in the first three months (from 1:1,280 to 1:2,560); they then declined to levels ranging from 1:80 to 1:320. The specificity of these antibodies for HIV-1 was confirmed by a competition ELISA of all serum samples, using human serum which was strongly positive for all the viral proteins, as well as purified monospecific antibodies raised in sheep against p24, gp41 and gp120 peptides (Biochrom). All specimens were tested at a dilution of 1 in 200 for HIV-1 p24 antigen to avoid the high background of undiluted rabbit sera. Nevertheless, p24 protein was present after two weeks (~ 250 pg ml⁻¹), reached high levels in the first three months (up to 1,500 pg ml⁻¹), and was still present over one year after infection

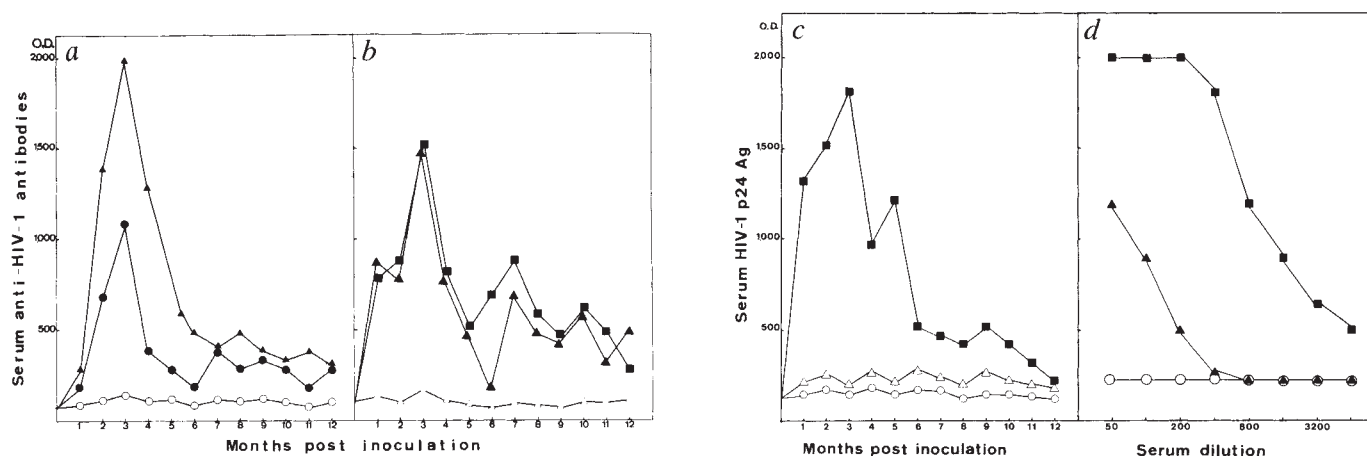


Fig. 1 *a*, HIV-1 specific IgG antibody responses of HTLV-III_B treated animals (SA ▲; SB ●; and control C3 ○). *b*, Responses of H9/HTLV-III_B inoculated rabbits (G1 ▲; G2 ■; and control C1 △). *c*, HIV-1 p24 antigen in sera pooled from HTLV-III_B inoculated rabbits (SA+SB ■) and control rabbits (C1+C2+C3 ○). The specificity of detected antigen was verified in all samples by neutralization test with human polyvalent anti-HIV-1 serum (△). *d*, SA+SB pooled sera, at different dilutions, were incubated with human polyvalent anti-HIV-1 serum (▲) or with a human negative serum (■). Rabbit negative control (○). HIV-1 antibody and p24Ag detection was performed as described in Table 1 legend. Results are the mean values of weekly determinations.

Table 1 HIV-1 infection of rabbits

Rabbit	Date of inoculation	Inoculum	Detection* of antibody (days)	Detection* of p24Ag (days)	Virus isolation after inoculum (days)
G1	03-20-87	H9/HTLV-III _B	14-360	14-360	120, 240, 300
G2	03-20-87	H9/HTLV-III _B	14-360	14-360	120, 240, 300
G3	03-20-87	H9/HTLV-III _B	14-360	14-360	120, 240, 300
SA	03-20-87	HTLV-III _B	14-360	14-360	120, 240, 300
SB	03-20-87	HTLV-III _B	14-360	14-360	120, 240, 300
SC	03-20-87	HTLV-III _B	14-360	14-360	ND
C1	03-20-87	H9	Negative	Negative	Negative
C2	03-20-87	H9	Negative	Negative	Negative
C3	03-20-87	Growth medium	Negative	Negative	Negative
CC1	10-14-87	Rabbit PBMC	14-180	14-180	210
CBX	02-26-88	H9/HIV-1	14-60	14-60	23, 49

The rabbits (HY/Cr) were purchased from Charles River Italia, with the exception of an outbred farm-raised animal (CBX), and were housed in isolation facilities. They all were negative for antibodies to HIV-1 by ELISA. The HTLV-III_B strain of HIV-1, the virus infected H9 cells and the H9 T-cell line used to inoculate rabbits were the generous gift of Dr R. C. Gallo. HTLV-III_B stock preparations were titred by end-point dilution in H9 cells and detection of reverse transcriptase activity in supernatants collected 10 days after infection. Aseptic peritonitis was induced by i.p. injection of 5 ml of thioglycollate (4%). After 4-5 days, rabbits were injected i.p. with 5 ml of control preparations or those containing HIV-1 in the presence of IL-2 (0.5 µg ml⁻¹, Sigma). Serum samples were obtained weekly, coded and tested for HIV-1 specific antibodies using the Behring HIV ELISA microplates following a modified procedure. Briefly, 100 µl of test serum, diluted 1:50 in phosphate-buffered saline (PBS) with 0.05% Tween 20 (PBS-T), was added to appropriate wells. After 30 min incubation at 37 °C and extensive washing, 100 µl of 1:1,000 diluted peroxidase-conjugated, affinity-purified, goat anti-rabbit IgG serum (Cappel Laboratories) was added to each well, and the plates were incubated for 30 min at 37 °C. Microplates were washed 6 times with PBS-T, 100 µl of peroxidase substrate solution [0.01% O-phenylenediamine (Sigma) and 0.01% H₂O₂ in 50 mM citrate buffer, pH 4.9] was added to each well. After 10 min at room temperature the reaction was stopped with 100 µl of 0.5 M H₂SO₄ and the A₄₉₂ was read with a Flow Multiscan reader. HIV-1 p24 antigen (Ag) was detected in serum samples, diluted 200-fold using a sensitive modified ELISA. The Biotin microwells coated with IgG fractions obtained from human sera, which exhibited high titres of anti-HIV-1 antibodies, were used to capture viral antigens. Sera were incubated overnight at room temperature. The detector reagents were: a rabbit antibody to HIV-1, goat antibody to rabbit IgG conjugated with peroxidase and O-phenylenediamine solution (Abbott), or a cocktail of murine p24 monoclonal antibodies conjugated with biotin, streptavidin-peroxidase and tetramethylbenzidine (Cellular Products). Virus isolation was performed by co-cultivation of phytohaemagglutinin-activated rabbit PBMC with H9 cells (1:4). Cultures were propagated in RPMI medium 1640 medium supplemented with 20% heat-inactivated fetal calf serum, 1% 200 mM glutamine, 200 U ml⁻¹ of recombinant IL-2 and antibiotics. Ag, antigen. ND, not determined.

* Numbers indicate dates of seroconversion and of the last positive sample.

(Fig. 2c). The p24 antigen was neutralized by a human polyvalent anti-HIV-1 (Fig. 2d) and sheep anti-p24 antibodies.

Eight and ten months after inoculation, peripheral blood mononuclear cells (PBMC) from five inoculated rabbits (G1, G2, G3, SA and SB) and from the three controls (C1, C2 and C3) were pooled and grown in culture. A low but detectable level of p24 antigen was present in the cell supernatant after nine days: less than 1% of the cells could be stained with peroxidase-conjugated human anti-HIV-1 serum or with monoclonal antibodies to HIV-1 p24 protein¹¹. Budding particles and

extracellular virions with morphology typical of HIV-1 were visualized in rabbit mononuclear cells by electron microscopy (Fig. 2a). When the cells were co-cultivated with uninfected H9 cells, there were more p24-positive cells which had extracellular retroviral particles (Fig. 2a). Viral expression increased and reverse transcriptase activity (more than 180,000 c.p.m. per ml) and p24 antigen (more than 1,000 pg per ml) were detected in culture supernatants (Table 2). We then derived an H9 cell line which was chronically infected with the rabbit isolate (H9/HIV-1_{R1}). DNA was extracted from H9, H9/HTLV-III_B and

Table 2 HIV-1 isolation from rabbit peripheral blood mononuclear cells co-cultivated with H9 cells

Donor	Cultured cells	Collected days after inoculum	Cellular p24Ag (%)	Electron microscopy	Proviral DNA	Reverse transcriptase c.p.m. per ml
G1, G2, G3, SA, SB	PBMC	240	<1	Positive	ND	ND
	PBMC + H9	240	10	Positive	ND	180,000
	PBMC + H9	300	10	Positive	Positive	279,000
C1, C2, C3	PBMC + H9	240	0	ND	ND	900
	PBMC + H9	300	0	ND	ND	700
CBX	PBMC + H9	49	10	Positive	ND	ND
	H9		0	Negative	Negative	300
	H9/HTLV-III _B		30	Positive	Positive	350,000

HIV-1 p24Ag was detected by immunoperoxidase staining after incubation with anti-p24 monoclonal antibodies (Du Pont, BT-3) in fixed cells¹¹ and by the Du Pont p24 ELISA in supernatant. The reverse transcriptase activity was measured by incorporation of [³H]thymidine into poly(rA) on an oligo (dT) template¹⁸. DNA was extracted from approximately 5×10^6 cells according to standard procedures¹⁹. HIV-1 amplified sequences were identified by hybridization with synthetic probes¹². ND, not determined.

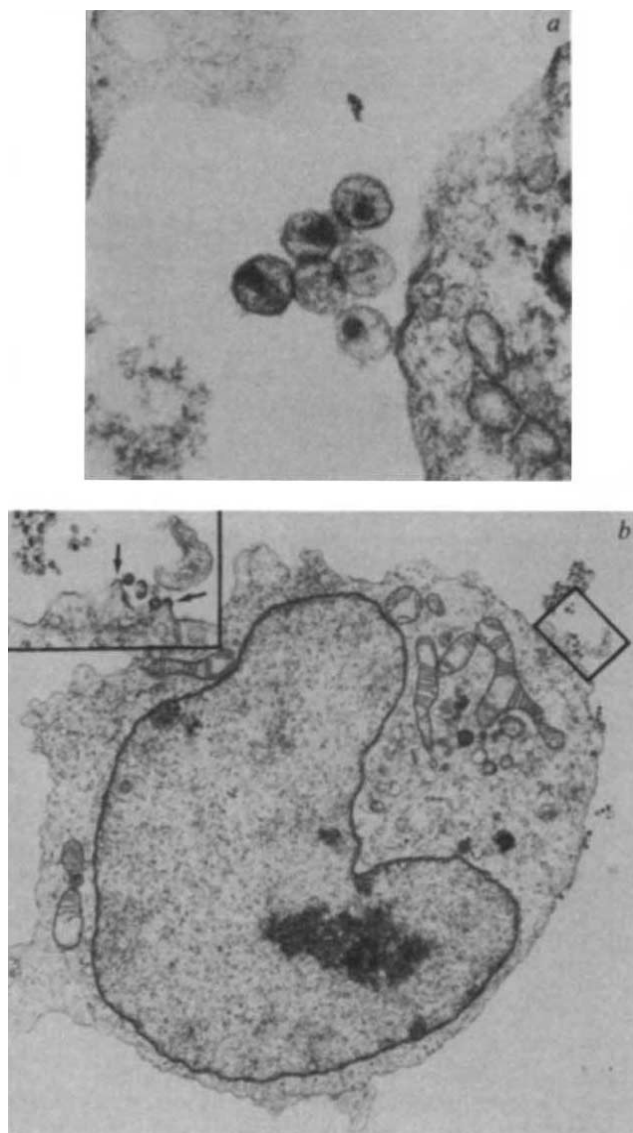


Fig. 2 Electron micrographs of *a*, extracellular immature particles with ring-shaped nucleoid and mature virions with conical core ($\times 105,000$) in rabbit mononuclear cells co-cultivated with H9 cells, and *b*, a rabbit mononuclear cell in a degenerative stage after 18 days in culture ($\times 4,900$) with budding retroviral particles and extracellular virions ($\times 31,500$).

H9/HIV-1_{R1} and sequences specific to HIV-1 were detected in preliminary experiments of gene amplification¹² (Table 2). The HIV-1_{R1} rabbit isolate was transmitted to another rabbit (CC1) by inoculation of PBMC pooled from 5 seropositive animals and re-isolated from PBMC up to 7 months after inoculation. HIV-1 infection was also induced in an outbred farm-raised rabbit (CBX) by i.p. injection of H9/HIV-1_{R1} cells (Table 1).

On the basis of results obtained so far it would seem that successful infection of rabbits requires inoculation into the peritoneum after hyperstimulation with thioglycollate. This treatment activates cells of the macrophage monocyte lineage. Although there is no evidence that the initially infected cells were macrophages, we rarely observed HIV-1 p24 antigen in adherent cells before co-cultivation with H9 cells. Given that macrophages may also be a primary target for HIV-1 infection in man¹³, this result could be significant. The mechanism by which HIV-1 infects human macrophages is not completely understood, particularly with regard to the requirement for the CD4 receptor¹⁴⁻¹⁵. It is possible that rabbit lymphocytes and macrophages exhibit a receptor that mimics CD4, because a mouse anti-human CD4 monoclonal antibody cross-reacts with a subset of rabbit lymphocytes (manuscript in preparation).

Other studies have already shown that rabbits can be infected by HTLV-1 (ref. 16), and recently it was shown that two T-cell lines, one transformed by herpesvirus atelae and the other by HTLV-1, and one macrophage line, transformed by SV40, are susceptible to infection by HIV-1 *in vitro*¹⁷. We show the reproducible *in vivo* HIV-1 infection of rabbits with known amounts of infectious virus or virus-infected cells, but, as yet, we have observed no disease symptoms in inoculated animals. Moreover, the degree and kinetics of infection have not been accurately determined. These and other studies are in progress in an effort to develop the rabbit as a model for preclinical studies of candidate vaccines and antiviral drugs.

We thank Elio G. Rondanelli, Egidio Romero, Lorenzo Moretta and Carmine A. Romanzi for helpful discussion and Paolo Orsolini and Laura Soldini for experimental procedures, Flossie Wong-Staal for DNA assays, and Robert C. Gallo and Dani P. Bolognesi for advice. This work was supported in part by IRCCS "S. Matteo", CNR and AIRC grants.

Received 13 June; accepted 11 August 1988.

1. Barré-Sinoussi, F. *et al. Science* **220**, 868-871 (1983).
2. Gallo, R. C. *et al. Science* **224**, 500-503 (1984).
3. Matthews, T. J. *et al. AIDS Res. Hum. Retroviruses* **3**, 197-206 (1987).
4. Alter, H. J. *et al. Science* **226**, 549-552 (1984).
5. Gajdusek, D. C. *et al. Lancet* **i**, 55-56 (1985).
6. Fultz, P. N. *et al. J. Virol.* **58**, 116-124 (1986).
7. Fultz, P. N. *et al. J. infect. Dis.* **154**, 896-900 (1986).
8. Desrosiers, R. C. & Letvin, N. L. *Rev. Infect. Dis.* **9**, 438-446 (1987).
9. Morrow, W. J. W., Wharton, M., Lan, D. & Levy, J. A. *J. gen. Virol.* **68**, 2253-2257 (1987).

10. Popovic, M., Sarngadharan, M. G., Read, E. & Gallo, R. C. *Science* **224**, 497-500 (1984).
11. Cereda, P. M., Debiaggi, M. & Romero, E. *Microbiologica* **7**, 107-111 (1984).
12. Mullis, K. & Faloona, F. A. *Cold Spring Harb. Symp. quant. Biol.* **51**, 263-273 (1986).
13. Gartner, S. *et al. Science* **233**, 215-219 (1986).
14. Åsjö, B. *et al. Virology* **157**, 359-365 (1987).
15. Clapham, P. R. *et al. Virology* **158**, 44-51 (1987).
16. Kotani, S. *et al. Int. J. Cancer* **37**, 843-847 (1986).
17. Kulaga, H., Folks, T. M., Rutledge, R. & Kindt, T. J. *Proc. natn. Acad. Sci. U.S.A.* **85**, 4455-4459 (1988).
18. Poiesz, B. J. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 7415-7419 (1980).
19. Davis, L. G. *et al. in Basic Methods in Molecular Biology* 44-46 (Elsevier, New York, 1986).

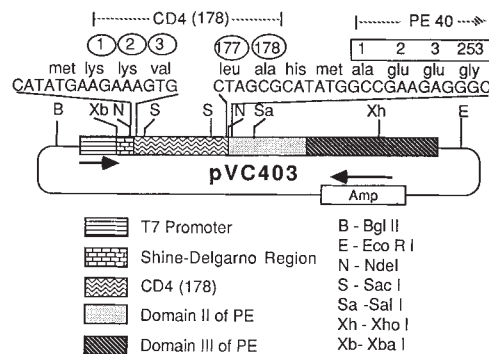


Fig. 1 Expression vector for CD4(178)-PE40. Expression plasmid pVC403 contains a fusion gene encoding the first 178 amino acids of mature CD4 [CD4(178)], based on the amino acid sequencing data reported in refs 11, 12] and amino acids 1-3 and 253-613 of PE (PE40, ref. 6). The gene is under control of a T7 late promoter²⁸. *E. coli* strain BL21 (λ DE3) carrying pVC403 was used to express the hybrid protein upon IPTG induction. The direction of transcription from the T7 promoter and for the β -lactamase gene is shown by the solid arrows. The circled numbers are the amino acids of CD4 and the boxed numbers are the amino acids of PE. The boundaries of the CD4(178) sequence and the start of the PE40 sequence are shown at the top. Plasmid pVC403 was constructed using standard recombinant DNA techniques. The detailed construction of the plasmid will be described elsewhere.

Selective killing of HIV-infected cells by recombinant human CD4-Pseudomonas exotoxin hybrid protein

Vijay K. Chaudhary*, Tamio Mizukami†, Thomas R. Fuerst†, David J. FitzGerald*, Bernard Moss†, Ira Pastan* & Edward A. Berger†

* Laboratory of Molecular Biology, National Cancer Institute, and
† Laboratory of Viral Diseases, National Institute of Allergy and Infectious Diseases; National Institutes of Health, Bethesda, Maryland 20892, USA

It is projected that in the absence of effective therapy, most individuals infected with human immunodeficiency virus (HIV) will develop acquired immune deficiency syndrome (AIDS) and ultimately succumb to a combination of opportunistic microbial infections, malignancies and direct pathogenic effects of the virus¹⁻³. Anti-viral agents, immunomodulators, and inhibitors of specific HIV functions are being tested as potential treatments to alleviate the high morbidity and mortality⁴. An alternative therapeutic concept involves the development of cytotoxic agents that are targeted to kill HIV-infected cells. Here we describe the purification and characterization of a recombinant protein produced in *Escherichia coli* that contains the HIV-binding portion of the human CD4 molecule linked to active regions of *Pseudomonas* exotoxin A. This hybrid protein displays selective toxicity toward cells expressing the HIV envelope glycoprotein and thus represents a promising novel therapeutic agent for the treatment of AIDS.

Targeted toxins represent a new approach to specific cytotoxic therapy. Using chemical coupling or gene fusion methodologies, hybrid molecules containing protein toxins linked to either antibodies or protein ligands have been shown to selectively kill cells expressing the corresponding surface antigens or receptor⁵⁻⁷. In applying this technology to the treatment of AIDS, we sought to replace the cell recognition domain (domain I) of *Pseudomonas* exotoxin A (ref. 6) with a portion of CD4, the T-lymphocyte surface molecule which serves as the HIV receptor^{8,9}. The proposed hybrid protein would be expected to bind selectively to HIV-infected cells expressing the viral envelope glycoprotein on their surface, thereby destroying these cells and limiting the spread of infection within the individual.

Recent studies indicate that the ability to bind gp120, the external subunit of the HIV envelope glycoprotein, is retained by soluble truncated CD4 molecules¹⁰⁻¹⁵ containing ~180 amino acids from the N-terminal end^{14,15}. Accordingly, we constructed an *E. coli* expression plasmid (pVC403) containing a chimaeric gene encoding the first 178 amino acids of mature CD4, plus amino acids 1-3 and 253-613 of *Pseudomonas* exotoxin A (PE) (Fig. 1). This segment of PE (designated PE40, ref. 6) lacks domain I but includes domains II and III that are responsible for translocation and ADP-ribosylation respectively. The hybrid protein, designated CD4(178)-PE40, was synthesized in *E. coli*, remained intracellular, and appeared to be primarily associated with inclusion bodies in the 100,000g pellet of sonicated spheroplasts. The denatured protein had the expected relative molecular mass (M_r) of ~60,000 (60K) and reacted with poly-

clonal antibodies to both PE and CD4 (see below).

A purification scheme involving denaturation of the insoluble protein with guanidine followed by rapid dilution was used. A substantially pure, monomeric form of CD4(178)-PE40 was obtained by chromatography of the renatured protein on a mono Q column (Fig. 2a). Analysis of each fraction by polyacrylamide gel electrophoresis (PAGE) under native conditions (data not shown) and in the presence of sodium dodecyl sulphate (SDS) (Fig. 2b) indicated that fraction 19 contained at least 80% pure monomeric hybrid protein with the expected M_r of 60K (lane 2) and that this protein reacted with rabbit polyclonal antibodies to both PE (lane 5) and CD4 (lane 8). This fraction displayed high ADP-ribosylation activity, 30-50% of which could be selectively immunoprecipitated by the anti-CD4 OKT4A monoclonal antibody (not shown). TSK-250 gel filtration analysis (data not shown) indicated that the hybrid protein in fraction 19 eluted as a symmetrical peak at the position expected for a protein of 60K, and SDS-PAGE demonstrated a single 60K band.

CD4(178)-PE40 was found to bind gp120 both in solution and on the cell surface. First, CD4(178)-PE40 was mixed with soluble [³⁵S]methionine-labelled gp120, and the immune complexes obtained with anti-PE serum were bound to protein A-agarose and resolved by SDS-PAGE. As shown in Fig. 3a, labelled gp120 was specifically co-precipitated along with CD4(178)-PE40. Second, the binding of CD4(178)-PE40 to cell-associated gp120 was established by immunofluorescence microscopy using polyclonal anti-PE antibodies. The HIV envelope glycoprotein was produced in CV-1 cells using a vaccinia-based expression system. It has been shown previously that both the external gp120 and the transmembrane gp41 subunits (derived by processing of the gp160 envelope precursor) are present on the surface of cells infected with recombinant vaccinia viruses containing the HIV envelope gene¹⁶⁻¹⁸. Figure 3b shows that CD4(178)-PE40 bound to cells that had been infected with a recombinant vaccinia virus encoding gp160, but not to cells infected with a control recombinant vaccinia virus. Additional specificity controls (not shown) indicated the fluorescence was not observed if either the CD4(178)-PE40 or the anti-PE antibodies were omitted from the reaction.

Binding of authentic PE to cells, followed by internalization and translocation to the cytoplasm, results in ADP-ribosylation

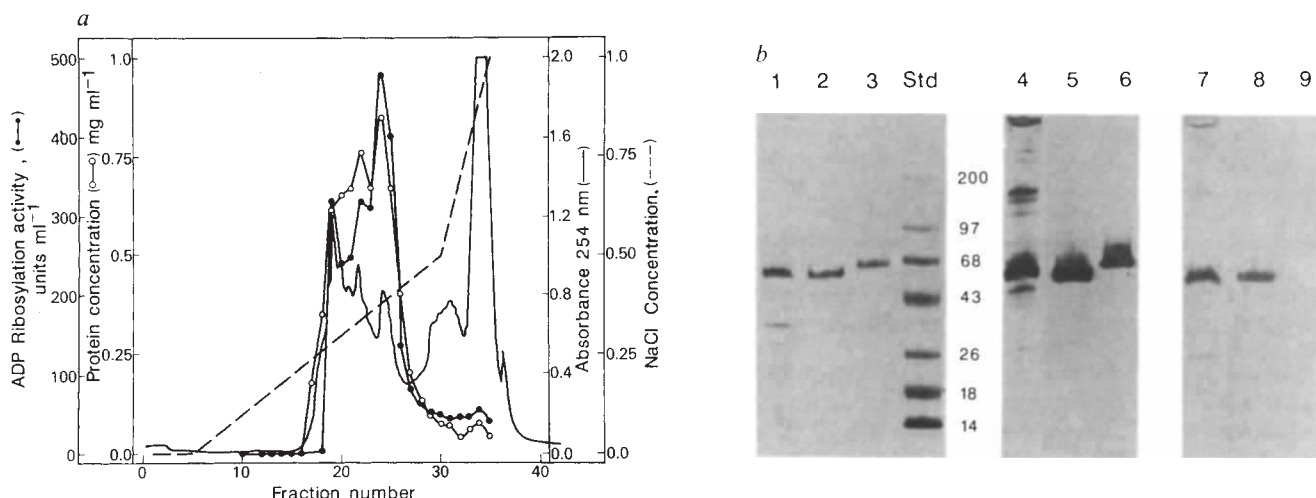


Fig. 2 Purification and characterization of CD4(178)-PE40. *a*, Mono Q column chromatography of renatured soluble CD4(178)-PE40. Renatured material (110 ml, 6 mg protein) was applied to a Mono Q column; proteins were eluted with a NaCl gradient and 1 ml fractions were collected. *b*, SDS-PAGE of samples at various stages of purification. Gels were either stained with Coomassie blue (lanes 1-3), or immunoblotted with polyclonal antibodies to PE (lanes 4-6) or CD4 (lanes 7-9). Lanes 1, 4 and 7, spheroplasts; lanes 2, 5 and 8, fraction 19 from Mono Q column; lanes 3, 6 and 9, authentic PE, 100 ng. Protein standards of known M_r are shown. M_r of authentic PE is 66K.

Methods. BL21 (λ DE3) carrying plasmid pVC403 was grown in LB medium at 37 °C with ampicillin (100 μ g ml⁻¹) and induced when the absorbance at 650 nm reached 0.6 with 1 mM isopropyl β -D-thiogalactoside (IPTG). The incubation was continued for 90 min, the cells harvested and fractionated into periplasm and spheroplasts²⁹. The spheroplasts were suspended in TE (50 mM Tris pH 8.0, 1 mM EDTA), sonicated three times at 100 watts for 30 s each and spun at 100,000g to isolate the supernatant (cytoplasm) and pellet (containing inclusion bodies). The various fractions were analysed by ADP-ribosylation assays³⁰, by SDS-PAGE³¹, and by immunoblotting with rabbit antibodies to CD4 (donated by R. Sweet, Smith Kline and French) and to PE. For partial purification of CD4(178)-PE40, a denaturation/renaturation procedure was employed. A 500 ml culture of BL21 (λ DE3) containing pVC403 was induced and the inclusion body pellet fraction prepared as described above. The pellet was suspended in 6.5 ml extraction buffer (guanidine HCl 7M, Tris HCl 0.1 M, pH 8.0, EDTA 1 mM and DTT 1 mM) and sonicated for 20 s three times. The suspension was stirred for 1 h in the cold and centrifuged at 100,000g for 15 min. The supernatant (6.5 ml) was added dropwise to 500 ml cold phosphate buffered saline with rapid stirring. After 48 h, a 110 ml aliquot was dialysed against buffer A (Tris-HCl 20 mM, pH 7.7) for 8 h, filtered through a 0.45 μ m filter and applied on a Mono Q column (HR 5/5) at 1 ml min⁻¹. The column was washed with 5 ml buffer A and developed with a 25 ml linear gradient (0-0.5 M NaCl). Fractions (1 ml) were collected and analysed for ADP-ribosylation and total protein, using Bradford reagent with bovine serum albumin as standard. ADP-ribosylation activity is expressed as units ml⁻¹; 1 unit is equal to the activity of 1 μ g of PE40. For SDS-PAGE, samples were boiled with Laemmli sample buffer³¹ and electrophoresed on 10-15% gradient gels (Phast gels, Pharmacia LKB).

of elongation factor 2 and the consequent inhibition of protein synthesis and cell death¹⁹. Two assay systems were employed to determine whether CD4(178)-PE40 is selectively internalized and translocated by cells expressing the HIV envelope glycoprotein, leading to inhibition of protein synthesis. First, the effects on cells infected with recombinant vaccinia viruses were examined. Figure 4a shows that protein synthesis in CV-1 cells infected with a recombinant vaccinia virus encoding gp160 was severely inhibited by CD4(178)-PE40 (ID_{50} of 27 ng ml⁻¹ under the conditions used). By contrast, cells infected with a control recombinant vaccinia virus were much less sensitive to the hybrid toxin (ID_{50} > 1,000 ng ml⁻¹). When authentic PE was used, cells infected with the gp160-expressing or control vaccinia viruses were equally sensitive (ID_{50} = 100 ng ml⁻¹); they were also equally insensitive to PE40 (ID_{50} > 1,000 ng ml⁻¹). We observed (not shown) that the anti-CD4 antibody OKT4A, which blocks the interaction between CD4 and gp120, neutralized the toxic activity of CD4(178)-PE40 on cells expressing the HIV envelope glycoprotein, but had no effect on the toxicity caused by authentic PE. By contrast, the anti-CD4 monoclonal antibody OKT4, which does not block the CD4/gp120 interaction, had no effect on the activity of either protein. We conclude that expression of the HIV envelope glycoprotein rendered cells very sensitive to the hybrid toxin and that the specificity was conferred by the CD4 moiety.

As a second system to evaluate the selectivity and effectiveness of CD4(178)-PE40, an uninfected human lymphocyte cell line (A3.01) and a daughter cell line (8E5) that is chronically infected with HIV were tested as targets. The 8E5 cells are especially suitable for experimental studies as they (1) contain a single

integrated viral genome, (2) constitutively synthesize HIV proteins, including gp120, (3) form syncytia when mixed with CD4-bearing cells, and (4) produce budding particles^{20,21}. Addition of CD4(178)-PE40 to 8E5 cells led to inhibition of protein synthesis: the ID_{50} of 100 ng ml⁻¹ under the conditions used indicated that the HIV-infected cells were highly sensitive to the hybrid toxin (Fig. 4b). By contrast, protein synthesis in the parental A3.01 cells was completely resistant to CD4(178)-PE40. Both cell lines were moderately sensitive to authentic PE (ID_{50} = 500 ng ml⁻¹) and unaffected by PE40.

In evaluating the therapeutic potential of a hybrid toxin, effects on cells other than the desired targets must be considered. B-lymphocytes and macrophages might, for instance, be affected by the chimaeric toxin as during normal T-lymphocyte function, CD4 is believed to interact with class II major histocompatibility (MHC) molecules on the surface of antigen-presenting cells²²⁻²⁵. In preliminary experiments, however, CD4(178)-PE40 did not inhibit protein synthesis in Raji cells, a B-cell line which expresses relatively large amounts of MHC class II molecules (data not shown). This result is consistent with the report that soluble CD4 shows no inhibitory effect on *in vitro* cell:cell interactions mediated by CD4 and MHC class II molecules¹². But the potential of the hybrid toxin to interfere with functions mediated by MHC class II molecules requires further exploration.

Several recent reports show that soluble CD4 alone can inhibit HIV infectivity in cell culture, presumably by competing for binding of the virus to normal cell-associated CD4¹⁰⁻¹⁴. The hybrid toxin described here represents an alternative anti-viral application of a soluble CD4 derivative, namely targeted killing

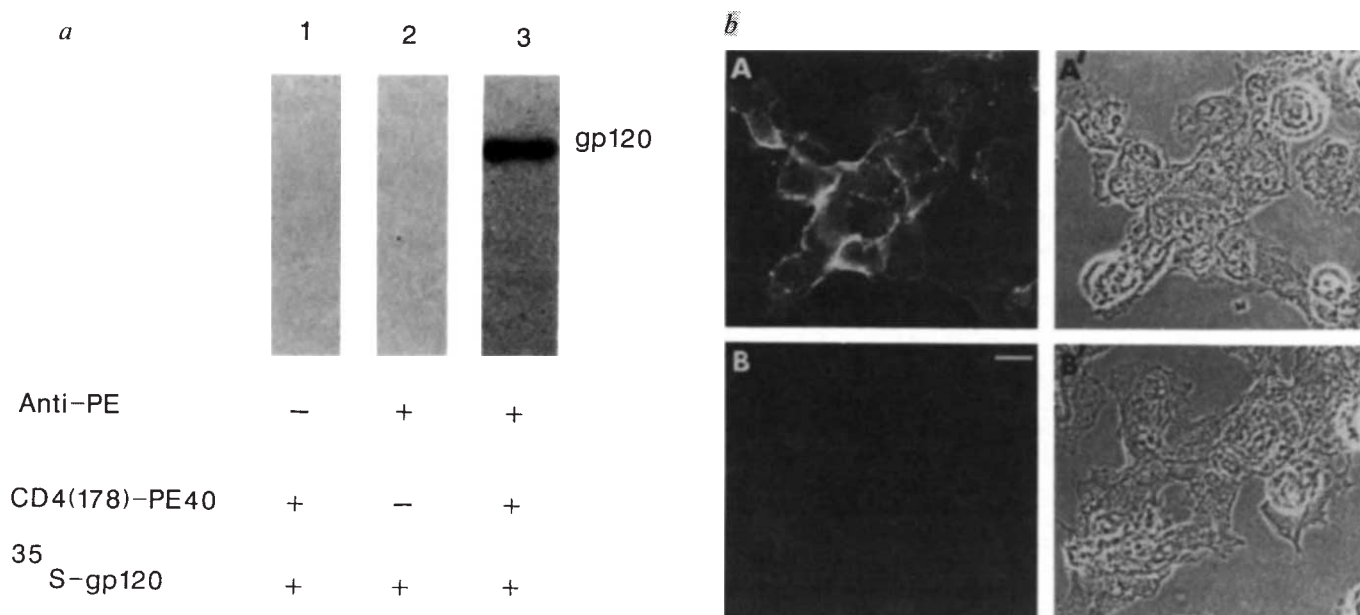
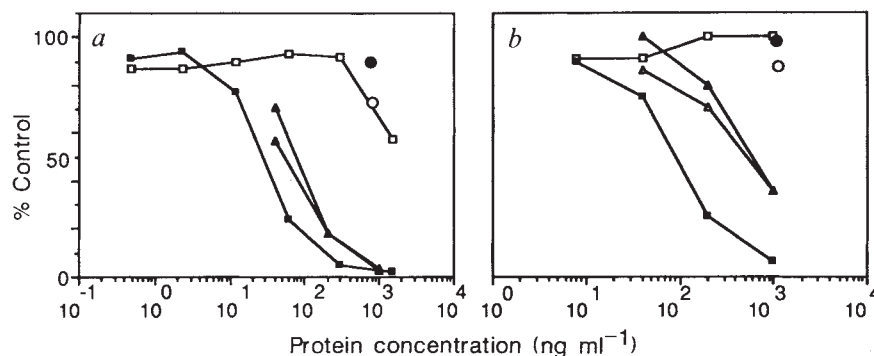


Fig. 3 Binding of CD4(178)-PE40 to the HIV envelope glycoprotein. *a*, Co-precipitation of recombinant gp120 with CD4(178)-PE40. Media containing ³⁵S-labelled gp120 was preincubated with CD4(178)-PE40 and the complexes precipitated with anti-PE plus Protein A-agarose (lane 3). For controls, either the anti-PE (lane 1) or the CD4(178)-PE40 (lane 2) were omitted. *b*, Specific binding of CD4(178)-PE40 to HIV envelope protein at the cell surface. CD4(178)-PE40 was added to CV1 cells infected with either a vaccinia virus recombinant containing the HIV-1 gp160 gene (*A* and *A'*), or a control vaccinia recombinant (*B* and *B'*). Surface-bound CD4(178)-PE40 was detected using rabbit polyclonal antibody to PE followed by rhodamine-conjugated goat anti-rabbit IgG. *A* and *B* are fluorescence micrographs, whereas *A'* and *B'* are the corresponding phase contrast in micrographs. Bar indicates 20 μ m at 350 $\times$ magnification.

Methods. *a*, Medium containing [³⁵S]methionine-labelled gp120 was obtained¹⁵ using a vaccinia/bacteriophage T7 hybrid expression system³². Medium (5 μ l) was preincubated with 90 μ l of crude renatured CD4(178)-PE40 for 4 h at 4 $^{\circ}$ C. Anti-PE antiserum (2 μ l) was added, and after overnight incubation at 4 $^{\circ}$ C the immune complexes were precipitated with 10 μ l protein A-agarose (Calbiochem). The immunoprecipitates were analysed by SDS-PAGE on 10% gels³¹ and the protein bands visualized by fluorography. *b*, Confluent monolayers of CV-1 cells were infected with vPE16, a recombinant vaccinia virus encoding gp160 (HIV-1, IIIB isolate) under control of the vaccinia 7.5K promoter. The multiplicity of infection was 1.5. The vPE16 recombinant virus produces a higher level of envelope protein than the previously described^{16,18} vSC25 (P. Earl, personal communication). As a control, we used vTF7-3, a vaccinia recombinant encoding the bacteriophage T7 RNA polymerase driven by the vaccinia 7.5K promoter³². Ten hours post-infection, CD4(178)-PE40 (fraction 19 of the Mono Q column, see Fig. 2) was added to a concentration of 50 μ g ml⁻¹ in PBS plus 0.2% (w/v) BSA. After 1 h at 4 $^{\circ}$ C, the dishes were rinsed, incubated with anti-PE antiserum [1:500 in PBS plus 0.2% (w/v) BSA] for 1 h at 4 $^{\circ}$ C, and incubated with affinity-purified goat anti-rabbit IgG conjugated to rhodamine. Cells were fixed in formaldehyde, mounted and photographed.

Fig. 4 Selective cytotoxicity of CD4(178)-PE40 for cells expressing the HIV envelope glycoprotein. *a*, Cells expressing the HIV envelope glycoprotein encoded by a recombinant vaccinia virus. Closed symbols represent cells infected with vPE16 encoding gp160, whereas open symbols represent cells infected with a control vaccinia recombinant, vTF7-3 (see legend to Fig. 3). The toxin preparations used were: ($\blacktriangle$, $\triangle$) PE; ($\bullet$, $\circ$) PE40; ($\blacksquare$, $\square$) CD4(178)-PE40 (fraction 19 of the Mono Q column shown in Fig. 2). One nM toxin corresponds to 66 ng ml⁻¹ for PE, 40 ng ml⁻¹ for PE40, and 60 ng ml⁻¹ for CD4(178)-PE40. *b*, Cells chronically infected with HIV. Closed symbols represent the 8E5 human T-cell line which continuously produces virions that are non-infectious as a result of a premature chain termination mutation in the reverse transcriptase gene^{20,21}. Open symbols represent the parental non-infected A3.01 cell line. The toxin preparations used were: ($\blacktriangle$, $\triangle$) PE, ($\bullet$, $\circ$) PE40, ($\blacksquare$, $\square$) CD4(178)-PE40.

Methods. *a*, Duplicate assays were performed on CV-1 cells in 24-well plates (2×10^5 cells per well). Cells were infected with the indicated recombinant vaccinia viruses at a multiplicity of infection of 20. After 9 h, toxin preparations were added, and 4 h later the cells were pulse-labelled for 1 h with [³⁵S]methionine. Radioactivity in trichloroacetic acid precipitates was counted and expressed as % control incorporation (no toxin added). *b*, Duplicate wells of 24-well plates were seeded with 2×10^5 cells of the indicated cell line. Toxin preparations were added, and after 17.5 h, the cells were pulse-labelled and processed as described above.



of HIV-infected cells. We anticipate that CD4(178)-PE40 will prove effective against cells infected with diverse isolates of HIV-1 as well as HIV-2, because the envelope proteins of all these viruses retain binding specificity for CD4 (ref. 26) despite extensive sequence variation²⁷. The fusion toxin may, therefore, have significant advantages over monoclonal antibody-toxin

conjugates which could possibly be developed for similar purposes. The potential therapeutic value of this novel reagent in the treatment of AIDS warrants further study.

We thank P. Earl for the vaccinia virus recombinants containing HIV envelope genes, T. Folks for the A3.01 and 8E5 cell lines, and R. Sweet for the polyclonal antibodies against CD4.

We are grateful to M. C. Willingham for help with the fluorescence microscopy, S. Adhya for scientific advice, N. Cooper, B. Lovelace and M. Gallo for technical assistance, and A. Gaddis for preparation of the manuscript. T.M. was supported by a fellowship from Kyowa Hakko Kogyo Co., Ltd.

Note added in proof: When treatment of 8E5 cells with CD4(178)-PE40 was extended to five days its ID50 was 5 ng ml^{-1} .

Received 24 June; accepted 8 August 1988.

- Anderson, R. M. & May, R. M. *Nature* **333**, 514–518 (1988).
- Levy, J. A. *Nature* **333**, 519–522 (1988).
- Fauci, A. S. *Science* **239**, 617–622 (1988).
- Mitsuya, H. & Broder, S. *Nature* **325**, 773–778 (1987).
- Pastan, I., Willingham, M. C. & Fitzgerald, D. J. *Cell* **47**, 641–648 (1986).
- Chaudhary, V. K., Fitzgerald, D. J., Adhya, S. & Pastan, I. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4538–4542 (1987).
- Lorberbaum-Galski, H., Fitzgerald, D., Chaudhary, V., Adhya, S. & Pastan, I. *Proc. natn. Acad. Sci. U.S.A.* **85**, 1922–1926 (1988).
- Littman, D. R. A. *Rev. Immun.* **5**, 561–584 (1987).
- Sattentau, Q. J. & Weiss, R. A. *Cell* **52**, 631–633 (1988).
- Smith, D. H. *et al. Science* **238**, 1704–1707 (1987).
- Fisher, R. A. *et al. Nature* **331**, 76–78 (1988).
- Hussey, R. E. *et al. Nature* **331**, 78–81 (1988).
- Deen, K. C. *et al. Nature* **331**, 82–84 (1988).
- Trautnecker, A., Luke, W. & Karjalainen, K. *Nature* **331**, 84–86 (1988).
- Berger, E. A., Fuerst, T. R. & Moss, B. *Proc. natn. Acad. Sci. U.S.A.* **85**, 2357–2361 (1988).
- Chakrabarti, S., Robert-Guroff, M., Wong-Staal, F., Gallo, R. C. & Moss, B. *Nature* **320**, 535–537 (1986).
- Hu, S., Kosowski, S. G. & Dalrymple, J. M. *Nature* **320**, 537–539 (1986).
- Lifson, J. D. *et al. Nature* **323**, 725–728 (1986).
- Middlebrook, J. L. & Dorland, R. B. *Microbiol. Rev.* **48**, 199–221 (1984).
- Folks, T. M. *et al. J. exp. Med.* **164**, 280–290 (1986).
- Gendelman, H. E. *et al. Virology* **160**, 323–329 (1987).
- Biddison W. E. & Shaw, S. *Diagnostic Immun.* **1**, 112–115 (1983).
- Romain, P. L., Schlossman, S. F. & Reinherz, E. J. *Immun.* **133**, 1093–1100 (1984).
- Gay, D. *et al. Nature* **328**, 626–629 (1987).
- Doyle, C. & Strominger, J. L. *Nature* **330**, 256–259 (1987).
- Sattentau, *et al. AIDS* **2**, 101–105 (1988).
- Coffin, J. M. *Cell* **46**, 1–4 (1986).
- Studier, F. W. & Moffatt, B. A. *J. molec. Biol.* **189**, 113–130 (1986).
- Chaudhary, V. K., Xu, Y., Fitzgerald, D., Adhya, S. & Pastan, I. *Proc. natn. Acad. Sci. U.S.A.* **85**, 2939–2943 (1988).
- Collier, R. J. & Kandel, J. J. *biol. Chem.* **246**, 1496–1503 (1971).
- Laemmli, U. K. *Nature* **227**, 680–685 (1970).
- Fuerst, T. R., Niles, E. G., Studier, F. W. & Moss, B. *Proc. natn. Acad. Sci. U.S.A.* **83**, 8122–8126 (1986).

Activation *in vitro* of sequence-specific DNA binding by a human regulatory factor

Jeffrey S. Larson, Thomas J. Schuetz & Robert E. Kingston*

Department of Genetics, Harvard Medical School and Department of Molecular Biology, Wellman 10, Massachusetts General Hospital, Boston, Massachusetts 02114, USA

The human heat-shock factor (HSF) regulates heat-shock genes in response to elevated temperature^{1–6}. When human cells are heated to 43 °C, HSF is modified post-translationally from a form that does not bind DNA to a form that binds to a specific sequence (the heat-shock element, HSE^{7,8}) found upstream of heat-shock genes⁶. To investigate the transduction of the heat signal to HSF, and more generally, how mammalian cells respond at the molecular level to environmental stimuli, we have developed a cell-free system that exhibits heat-induced activation of human HSF *in vitro*. Comparison of HSF activation *in vitro* and in intact cells suggests that the response of human cells to heat shock involves at least two steps. First, an ATP-independent, heat-induced alteration of HSF allows it to bind the HSE; the temperature at which activation occurs *in vitro* implies that a human factor directly senses temperature. Second, HSF is phosphorylated. It is possible that similar multi-step activation mechanisms play a role in the response of eukaryotic cells to a variety of environmental stimuli, and that

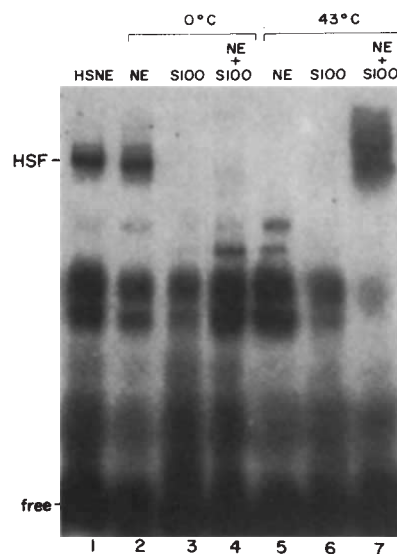


Fig. 1 Heat-inducible *in vitro* activation of DNA binding as determined by the gel mobility shift assay¹⁸. Nuclear and cytoplasmic extracts from 4 l of HeLa cells growing at 37 °C (denoted NE and S100, respectively), and nuclear extracts from 1 l of HeLa cells immediately after a 1 h incubation at 43 °C (HSNE) were prepared as described by¹⁷. Where indicated, 5.0 µg of protein of the NE (lanes 2 and 5), 8.5 µg S100 (lanes 3 and 6), or both (lanes 4 and 7) were incubated for 1 h at 0 °C (lanes 2–4) or 43 °C (lanes 5–7). Lane 1, 2.0 µg HSNE. The indicated extracts were then incubated at 30 °C for 30 min with a ³²P-labelled synthetic oligonucleotide containing the HSE (bases –115 to –80 of the human hsp70 promoter; ref. 6) and then electrophoresed on a non-denaturing polyacrylamide gel. HSF, heat-shock factor; free, unbound DNA. **Methods.** Extracts were prepared¹⁷ and dialysed into buffer D (20 mM HEPES, pH 7.9, 20% (v/v) glycerol, 100 mM KCl, 0.2 mM EDTA, 0.5 mM phenylmethylsulphonyl fluoride, 0.5 mM dithiothreitol) and were incubated at the indicated temperature without any addition. Cytoplasmic extracts ranged in concentration from 8.5 to 20 mg ml^{–1}, and were heated without dilution. More dilute incubation conditions dramatically reduce the degree of activation of HSF. Binding reaction mixtures (20 µl) contained 12 mM HEPES, pH 7.9, 12% (v/v) glycerol, 60 mM KCl, 0.12 mM EDTA, 0.3 mM phenylmethylsulphonyl fluoride, 0.3 mM dithiothreitol, 2 mM MgCl₂, 100 µg of poly(dI–dC) ml^{–1}, ~0.2 ng of the labelled DNA, and the indicated extract. These were then subjected to electrophoresis on a 4% non-denaturing polyacrylamide gel.

these mechanisms evolved to increase the range and flexibility of the response.

Coordinate regulation of mammalian genes often involves modification of a pre-existing transcription factor to a form that can bind to the regulatory element of a promoter^{6,9–15}. We report here the development of an *in vitro* system that faithfully mimics heat-induced binding of HSF to the HSE. We refer to HSF that can bind to the HSE as active HSF. Nuclear extracts from HeLa cells growing at 43 °C contain significantly more active HSF than extracts at 37 °C (refs 6, 16; Fig. 1, lanes 1 and 2). Nuclear and cytoplasmic extracts were prepared¹⁷ from HeLa cells growing at 37 °C, incubated at either 0 °C or 43 °C for one hour and the level of active HSF determined using band retention¹⁸ (Fig. 1). A 36-base pair double-stranded oligonucleotide containing an HSE—bases –80 to –115 of the human promoter of the gene encoding the heat-shock protein (hsp) of relative molecular mass 70,000 (70 K); see Fig. 2b—was used as labelled probe. The nuclear extract contained no binding activity either before or after heating, whereas the cytoplasmic extract had significant levels of binding activity only after incubation at 43 °C (Fig. 1, lanes 3 and 6).

The cytoplasmic heat-inducible binding activity formed a single retained band that co-migrated with the lower of two retained bands obtained with nuclear extract from heat-shocked

* To whom correspondence should be addressed

Fig. 2 Characterization of the sequence specificity of the heat-inducible binding activity present in the S100 extract. *a*, Competition of the heat-induced band with fragments containing the heat-shock element and nonspecific fragments. Gel mobility shift assays contained 8.5 μ g of the S100 protein not incubated (lane 1) or incubated at 43 °C for 1 h (lanes 2–12) before the DNA binding reaction. In addition to S100 protein and radioactive probe, binding reactions also contained non-radioactive competitor DNA (in a five or tenfold molar excess, as indicated): lanes 1 and 2, no competitor; lanes 3 and 4, bases –84 to +5 of the human hsp70 gene (no heat-shock element); lanes 5 and 6, bases –148 to –74 of the human hsp70 gene (heat-shock element centred at base –100); lanes 7 and 8, a 310-basepair fragment of *Drosophila melanogaster* heat-shock gene containing three heat-shock elements (*Xho*I [base –200] to *Eco*RI [base +110], of plasmid pSPS-HS-9, ref. 28), derived from cloned fragment 232 (ref. 29); lanes 9 and 10, a 39-base pair double-stranded synthetic oligonucleotide containing the CAAT sequence (top strand, 5'-CA CCG TCG ATT TCC CTT CTG AGC CAA TCA CCG AGC TCG A-3'); and lanes 11 and 12, the 36-base pair double-stranded synthetic oligonucleotide used for probe. *b*, Methylation interference analysis of binding to the heat-shock element by heated cytoplasmic extract (IN VITRO, lanes 1, 2, 3, 7, 8 and 9 and extract from heat-shocked HeLa cells (IN VIVO, lanes 4, 5, 6, 10, 11 and 12). Radioactive label was on the top strand (lanes 1 to 6) or on the bottom strand (lanes 7 to 12). Methylation-specific cleavage of DNA is represented as follows: lanes 1, 6, 7 and 12, DNA not incubated with extract (L: G. ladder); lanes 3, 4, 9 and 10, DNA incubated with extract and eluted from shifted bands (B: bound complex DNA); and lanes 2, 5, 8 and 11, uncomplexed DNA from the retention gel (F: free, unbound DNA). The sequence of the human hsp70 promoter fragment containing the heat-shock element used in this study is at the bottom of the figure. The heat-shock consensus sequence is represented by large letters (C-GAA-TTC-G). Solid arrows indicate the G residues under-represented in the bound DNA and the open arrow indicates a G residue partially under-represented in the bound DNA.

Methods. *a*, S100 extracts were prepared and DNA incubations and electrophoresis were performed as described in the Fig. 1 legend, except the binding reactions also contained unlabelled competitor DNA where indicated. DNA concentrations were estimated by agarose gel electrophoresis. *b*, Cytoplasmic extract was heated at 43 °C for 1.5 h. Binding reactions were performed as described in Fig. 1: 120 μ g of heated cytoplasmic extract or 16 μ g HSNE were incubated with 20 ng of a partially methylated, labelled fragment containing bases –148 to –74 of the promoter. Bound and free complexes were separated on a non-denaturing gel, eluted, and analysed as described²⁰.

cells (Fig. 1, compare lanes 1 and 6). We confirmed that this binding was sequence specific by competition with unlabelled DNA fragments and by methylation interference^{19,20}. All unlabelled fragments that contained an HSE competed for binding; there was no competition by fragments that did not contain the HSE (Fig. 2*a*). Methylation of five G residues in or adjacent to the HSE inhibited binding of HSF prepared either from intact heat-shocked HeLa cells or activated *in vitro* (Fig. 2*b*, compare lanes 3 and 9 with 4 and 10). These data demonstrate that the binding activity induced *in vitro* has the same sequence specificity as that seen after induction of HSF in whole cells⁶.

There is close correlation between the optimal temperature for *in vitro* activation and the temperature at which HeLa cells undergo the heat-shock response. There is very slight induction of the binding ability at 37 °C *in vitro*, increased induction at 40 °C and 43 °C, none at 50 °C (Fig. 3*a*). Binding ability is stable after induction at 43 °C, further incubation on ice, at 30 °C or at 37 °C for up to one hour does not cause loss of activity (Fig. 3*a*). Maximal induction of active HSF occurs at slower rate than *in vivo* (Fig. 3*b*). Activation of HSF in fractions from ion-exchange columns is unaffected by ATP addition, implying that the activation process is ATP-independent (data not shown).

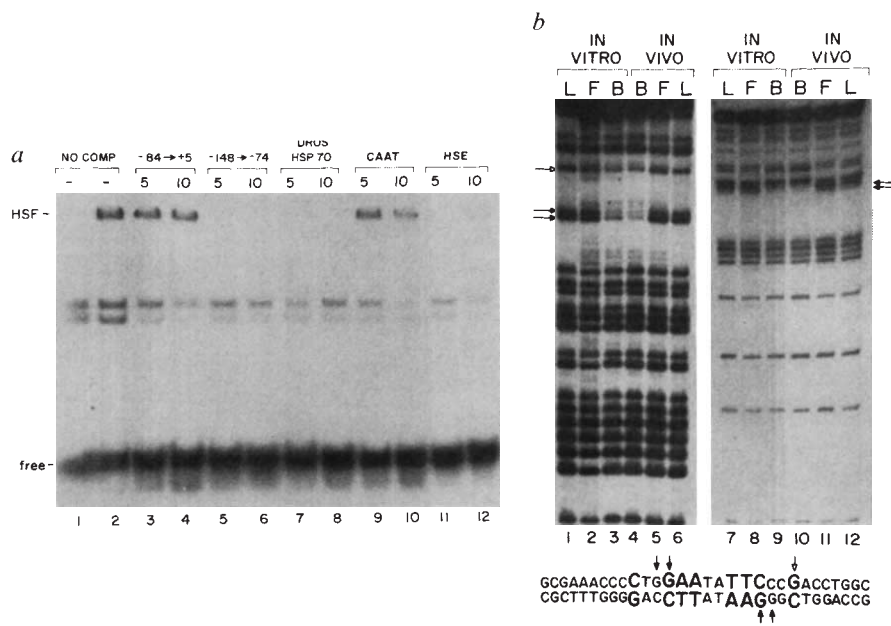
The specificity of HSE-binding is identical after *in vitro* activation and after HSF is isolated from heat-shocked HeLa cells (Fig. 2). Activation is efficient; induction of a cytoplasmic extract *in vitro* results in 12.2 units of binding activity per 10⁶ cells, whereas nuclei of heat-shocked HeLa cells contain 11.8 units of binding activity per 10⁶ cells. Most striking is the temperature at which activation occurs *in vitro*; there is little activation at 37 °C, the normal growth temperature of human cells, but extensive activation at 43 °C, the temperature at which human cells undergo heat shock. We cannot prove that the mechanism of activation *in vitro* is the same as *in vivo*, but the characteristics

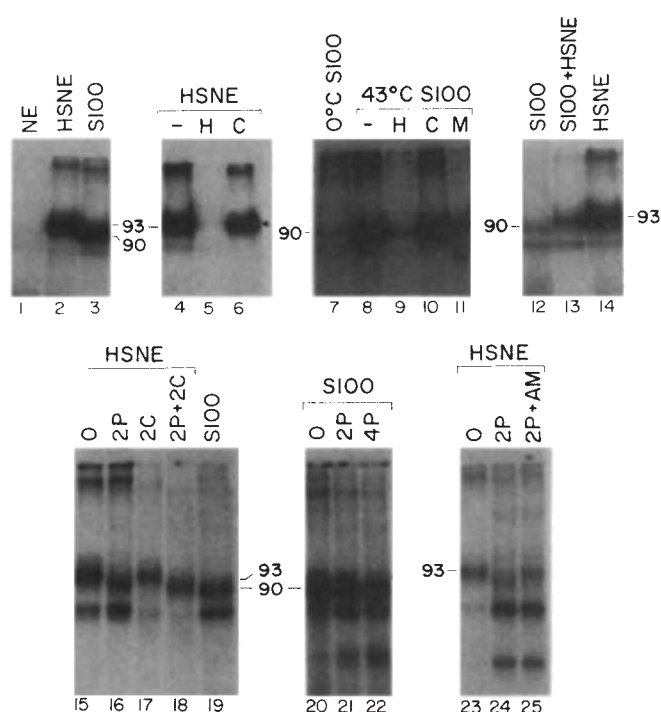
of *in vitro* activation imply that a factor exists in this extract, and hence in human cells, that can sense temperature.

How does such a factor sense heat? Heated cells are presumed to contain increased levels of denatured protein, and it has been suggested that this triggers activation of HSF through a protein degradation pathway^{21–24}. This model predicts that activation occurs in a temperature-dependent fashion in the *in vitro* system because there is more denatured protein at 43 °C than at 37 °C. However, addition of boiled bovine serum albumin (BSA) to the cytoplasmic extract did not change the rate of induction of HSF at either 37 °C or 43 °C (Fig. 3*c*) nor did it change the extent of induction (not shown). Addition of boiled cytoplasmic extract also had very little effect on activation (not shown). These results are consistent with a model in which the binding ability of HSF is altered by a specific factor, possibly HSF itself, that directly senses heat.

We performed ultraviolet (UV) cross-linking studies to determine the apparent size of HSF²⁵. Extracts were incubated with a radiolabelled and bromo-deoxyuridine-substituted HSE. Complexes were then separated on an agarose gel and protein was cross-linked to the labelled probe by irradiation with UV light. SDS-polyacrylamide gel electrophoresis of the resulting complexes from nuclear extracts revealed a 93K band (Fig. 4). This band has characteristics of HSF: it is not present in control (non-shocked) nuclear extracts (Fig. 4, lane 1); it is present in the HSF-HSE complex identified on mobility shift gels; and its formation is efficiently competed by an oligonucleotide that contains an HSE, but not by a control oligonucleotide specifying the CCAAT element (Fig. 4, lanes 4 to 6).

Experiments with the heated cytoplasmic extract showed, surprisingly, that HSF activated *in vitro* runs much faster (90K) than HSF from intact cells (Fig. 4, lanes 2 and 3). The 90K band is not present in cytoplasmic extract that has not been





Methods. UV-cross linking was performed exactly as described²⁵ except that binding reactions contained 6 ng [α^{32} P]CTP, bromodeoxyuridine-substituted probe, and 30 μ g extract (NE and HSNE) or 120 μ g extract (S100). For lane 13, 10 μ g HSNE and 80 μ g S100 were mixed, incubated for 1 h at 43 $^{\circ}$ C, and subjected to the cross-linking protocol. After separation of the HSF-HSE complex on a 1% low-melting-temperature agarose gel, the gel was irradiated for 7 min on a Fotodyne 3-3000 UV box, and the bands were visualized by autoradiography. The appropriate bands containing HSF were excised and analysed on a 10% SDS-polyacrylamide gel. The double-stranded oligonucleotide probe sequence was (top strand): 5'-GGT CGA CTG GAA TAT TCC GGA ATA TTC CGG GAT CCG AGC C-3'. No nuclease digestion was performed before running the protein gel. Phosphatase reactions were performed by incubating the appropriate extract (30 μ g HSNE or 120 μ g S100 in 30 μ l final volume) in buffer D (see Fig. 1 legend) with the indicated amount of phosphatase (Sigma, acid phosphatase from white potato, type VII) at 37 $^{\circ}$ C for one hour. S100 extracts were activated before phosphatase treatment by heating at 43 $^{\circ}$ C for 45 min. For lane 25, ammonium molybdate (10 μ M final concentration) was added before phosphatase. Two different preparations of potato acid phosphatase were used with identical results.

Methods. S100 extracts were prepared and activated and DNA incubation and electrophoresis were performed as described in the Fig. 1 legend, with the noted changes.

Yeast HSF increases in apparent molecular weight upon heat-shock of yeast cells without a change in DNA binding capabilities, and this increase seems to be caused by phosphorylation¹⁶. We treated both the HSF from heated HeLa cells and the *in vitro* activated HSF with phosphatase before cross linking the factor to DNA to see if phosphorylation was involved here.

Whereas treatment with calf intestinal phosphatase had very little effect (Fig. 4, lanes 15 and 17), treatment of HSF from heated cells with acid phosphatase or with both phosphatases increased the apparent mobility of HSF to form a diffuse band that extended from approximately 90K to 92K (lanes 16 and 18). This shift in mobility was inhibited by micromolar levels of ammonium molybdate, a specific inhibitor of acid phosphatase²⁶ (lanes 23 to 25). In contrast, there was no detectable effect of acid phosphatase on the mobility of the *in vitro* activated HSF (Fig. 4, lanes 20 to 22). Although phosphatase treatment did not change the mobility of HSF from heated cells to precisely that of *in vitro* activated HSF, these results indicate that HSF found in intact heated cells is more extensively phosphorylated than HSF found in non-heated cells.

These data therefore indicate that a change in binding ability is not the only alteration of HSF that occurs when cells are heated. We believe that HSF in human cells is modified in response to heat by at least two steps: an induction of sequence-specific binding ability followed by phosphorylation. Heating a cytoplasmic extract recreates only the first of these steps, resulting in a factor with lower apparent relative molecular mass. We propose that this phosphorylation plays a regulatory role in activation of HSF, although further studies are needed to prove this.

What advantage is there to this proposed multi-step modification of HSF? Factors that are fully activated by single step mechanisms must use that alteration as the only on-off switch for response to a regulatory effector. This limits the flexibility as well as the potential range of the response; for example, the lac repressor, a potent regulatory protein that alternates between forms that can and cannot bind a specific regulatory site, has at most a 1000-fold effect on β -galactosidase synthesis³⁷. The flexibility of the response could be increased by having each modification separately controlled by changes in the environmental effector. Similar multi-step processes may be involved in activation of other human genes in response to environmental signals.

We thank H. Prentice for making some of the extracts used in this work, S. Bray and J. Hirsh for suggesting the use of molybdate, D. Moore, F. Ausubel, I. C. A. Taylor, and R. Brent for comments on the manuscript, M. Cherry, L. Kizilay and J. Smith for the oligonucleotides, and R. Hyde for help with the manuscript. This work was supported by a grant from Hoechst AG.

Received 8 April; accepted 11 August 1988.

- Ritossa, F. M. *Experientia* **18**, 571-572 (1962).
- Nover, L. *Heat Shock Response of Eukaryotic Cells* (Springer, Berlin, 1984).
- Craig, E. A. *CRC Crit. Rev. Biochem.* **18**, 239-280 (1985).
- Pelham, H. R. B. *Trends Genet.* **1**, 31-35 (1985).
- Lindquist, S. A. *Rev. Biochem.* **55**, 1151-1191 (1986).
- Kingston, R. E., Schuetz, T. J. & Larin, Z. *Molec. cell. Biol.* **7**, 1530-1534 (1987).
- Pelham, H. R. B. & Bienz, M. *EMBO J.* **1**, 1473-1477 (1982).
- Mirault, M.-E., Southgate, R. & Delwart, E. *EMBO J.* **1**, 1279-1285 (1982).
- Yamamoto, K. R. A. *Rev. Genet.* **19**, 209-252 (1985).
- Sen, R. & Baltimore, D. *Cell* **47**, 921-928 (1986).
- Prywes, R. & Roeder, R. G. *Cell* **47**, 777-784 (1986).
- Hayes, T. E., Kitchen, A. M. & Cochran, B. H. *Proc. nat. Acad. Sci. U.S.A.* **84**, 1272-1276 (1987).
- Zimarino, V. & Wu, C. *Nature* **327**, 727-730 (1987).
- Seguin, C. & Hamer, D. H. *Science* **235**, 1383-1387 (1987).
- Baeuerle, P. A. & Baltimore, D. *Cell* **53**, 211-217 (1988).
- Sorger, P. K., Lewis, M. J. & Pelham, H. R. B. *Nature* **329**, 81-84 (1987).
- Dignam, J. D., Lebovitz, R. M. & Roeder, R. G. *Nucleic Acids Res.* **11**, 1475-1589 (1983).
- Fried, M. & Crothers, D. M. *Nucleic Acids Res.* **9**, 6505-6525 (1981).
- Siebenlist, U. & Gilbert, W. *Proc. nat. Acad. Sci. U.S.A.* **77**, 122-126 (1980).
- Gilman, M. Z., Wilson, R. N. & Weinberg, R. A. *Molec. cell. Biol.* **6**, 4305-4316 (1986).
- Anantham, J., Goldberg, A. L. & Voellmy, R. *Science* **232**, 522-524 (1986).
- Finley, D., Ciechanover, A. & Varshavsky, A. *Cell* **37**, 43-55 (1984).
- Goff, S. A. & Goldberg, A. L. *Cell* **41**, 587-595 (1985).
- Munro, S. & Pelham, H. *Nature* **317**, 477-478 (1985).
- Wu, C. *et al. Science* **238**, 1247-1253 (1987).
- Verjee, Z. H. M. *Eur. J. Biochem.* **9**, 439-444 (1969).
- Beckwith, J. & Zipser, D. *The lactose operon* (Cold Spring Harbor, New York, 1970).
- Wurm, F. M., Gwinn, K. A. & Kingston, R. E. *Proc. nat. Acad. Sci. U.S.A.* **83**, 5415-5418 (1986).
- Holmgren, R., Livak, K., Morimoto, R., Freund, R. & Meselson, M. *Cell* **18**, 1359-1370 (1979).

High resolution structure of the RNA duplex [U(U-A)₆A]₂

A. C. Dock-Bregeon*, B. Chevrier*, A. Podjarny*, D. Moras*, J. S. deBear†, G. R. Gough†, P. T. Gilham† & J. E. Johnson†

* Institut de Biologie Moléculaire et Cellulaire du CNRS, 15 rue René Descartes, 67084 Strasbourg, France

† Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907, USA

RNA is involved in many biological functions, ranging from information storage and transfer to the catalysis of reactions involving both nucleic acids and proteins. Previous crystallographic studies on RNA oligomeric chains provide only averaged structures¹ or information limited in resolution²⁻⁴. The oligomer [U(U-A)₆A]₂ was chosen for the study of protein-RNA interactions in viruses. Its size and base composition mimic portions of the genomic RNA in alfalfa mosaic virus that bind to the amino terminus of the viral subunit⁵. The actual sequence was designed to guarantee the formation of a single species of duplex and to facilitate the production of the pure oligomer in large quantities. The molecular structure, derived from the 2.25 Å resolution X-ray diffraction data, allows the most detailed analysis of an A-RNA helix reported to date. Two kinks are observed that divide the duplex into three blocks, each close to a canonical A-helical conformation. A few intermolecular hydrogen bonds involving 2'-hydroxyl groups stabilize this peculiar conformation of the RNA, which may be related to the temperature used for the crystallization (35 °C). The structure demonstrates both the plasticity of the RNA molecule and the role of the 2'-hydroxyl groups in intermolecular interactions.

The oligomer U(U-A)₆A was synthesized chemically in solution, using a modified version⁶ of the phosphotriester method

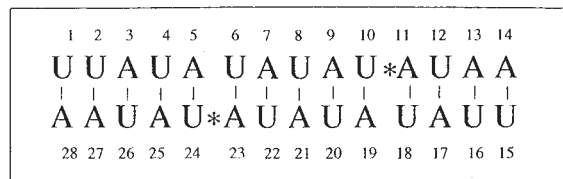


Fig. 1 Sequence of the duplex with asterisks at the kink points. For synthesis of the oligomer, 5'-O-dimethoxytrityl 2'-O-(*o*-nitrobenzyl) derivatives of uridine²⁴ and N⁶-benzoyladenine²⁵ were converted to their 3'-*p*-chlorophenyl phosphates²⁶, which were isolated as the barium salts⁶; at this point the uracil base was further protected by anisoylation²⁷. The barium salts were used to construct the dimer [(MeO)₂Tr]U^{an}(NBzl)]p(CIPh)A^{bz}(NBzl)-p(CIPh)(CN₂Et) which, after removal of appropriate groups and subsequent condensation, gave the protected tetramer UUAUA. One-third of this material was extended at its 5' end by condensation with the uridine monomer above, while N⁶, N⁶, O², O³-tetrabenzoyladenine was added to the 3' end of another third. The resulting blocks UUAUA, UUAUA, and UUAUA were then assembled on a 38 µmol scale to yield 180 mg of the protected tetradecanucleotide. Successive treatments with tetramethylguanidium pyridinealdoximate, ammonia, and longwave ultraviolet light at pH 3.5 (ref. 8) were used to remove the protecting groups. A portion of the product was purified by preparative HPLC, converted to its sodium salt, and lyophilized. The oligomer's base composition and molar absorptivity were determined by exonuclease digestion⁹; its sequence was confirmed by ribonuclease A degradation. The thermodynamic values for the duplex dissolved in 1.0 M NaCl, 0.01 M sodium phosphate, 0.1 mM EDTA, pH 7 are molar absorption coefficient (at 260 nm and 25 °C) = 99,000, T_m (at 3.8 µM strand concentration) = 43°, $\Delta H^\circ = -89.5 \text{ kcal mol}^{-1}$; $\Delta S^\circ = -258 \text{ eu}$; $\Delta G^\circ_{37} = -9.4 \text{ kcal mol}^{-1}$. Values calculated from nearest neighbour parameters²⁸: $\Delta H^\circ = -90 \text{ kcal mol}^{-1}$; $\Delta S^\circ = -262 \text{ eu}$; $\Delta G^\circ_{37} = -9.1 \text{ kcal mol}^{-1}$.

Table 1 Average torsion angles and helical parameters*

	U(U-A) ₆ A			tRNA ^{Asp}	RNA	RNA-DNA	A-DNA§	
	mean†	A11	U24	stems ⁴	fibre‡	hybrid ¹⁸	fibre	structures
Torsion angles								
α (P-O5')	-61	159	107	-74	-62	-69	-50	-73
β (O5'-C5')	169	177	191	183	180	175	172	178
γ (C5'-C4')	53	177	-142	74	47	55	41	62
δ (C4'-C3')	80	88	105	82	84	82	79	86
ϵ (C3'-O3')	-148	-114	-134	-147	-152	-151	-146	-154
ζ (O3'-P)	-79	-106	-78	-74	-74	-75	-78	-72
χ	-159	-178	-165	-166	-166	-162	-154	-158
Helical parameters, mean								
Twist angle		33.1		32.7	32.7	33.3	32.7	32.9
Rise per residue		2.79		2.38	2.81	2.52	2.56	3.00
Tilt		16.7		17.1	16.7	19.3	22.0	12.2
Propeller twist		18.6		8.5	13.8	11.8	6.0	11.5
Displacement		3.6		4.4	4.4	4.5	4.4	3.8
Groove width¶								
Minor		10.2		9.8	11.3	10.2	11.0	9.8
Major		3.7		4.1	4.1	3.2	2.3	7.0

* Helical parameters were calculated using the program HELIX in ref. 18, written by J. Rosenberg and R. E. Dickerson¹⁸. As the determination of a global helical axis was strongly affected by the kinks, the analysis was performed with the three domains (U1·A28 to A5·U24, A5·U24 to U10·A19 and U10·A19 to U15·A14) treated as separate helices.

† Not including A11 and U24, or A28 which is affected by end-effects.

‡ The parameters were measured or calculated from models of sequence U(U-A)₆A constructed with data from the fibre diffraction studies of RNA¹.

§ From the compilation by Shakked and Rabinovitch²³ of the helical parameters of d(GGTATACC), d(GGGGCCCC), d(CCGG), d(GGCCGGCC) (-8 °C), d(GGCCGGCC) and d(CCCCGGGG).

|| Distance of the C6-C8 vector from the helical axis.

¶ Defined as the phosphorus-to-phosphorus distance minus 5.8 Å to account for the Van der Waal's radii of the phosphate groups.

adapted for construction of oligoribonucleotides^{7,8} and its thermodynamic parameters were obtained from a least squares analysis of its melting curve⁹ (Fig. 1). The crystals of U(U-A)₆A were obtained by vapour diffusion at 35 °C from a solution of RNA (4 mM) containing magnesium (400 mM), methyl-2,2-pentanediol (35%) and sodium cacodylate (40 mM, pH 6.5). They belong to space group P2₁2₁2₁, with parameters $a = 34.11$ Å, $b = 44.61$ Å and $c = 49.11$ Å, and contain one duplex tetradecamer per asymmetric unit. The structure was solved by the molecular replacement method and refined to an *R*-factor of 13.4% at 2.25 Å resolution. A portion of the final 'omit' map is shown in Fig. 2.

The molecular structure presented in Fig. 2 is a right-handed A-type double helix. It differs significantly from the fibre-deduced model at two kink positions which divide the molecule in three domains of 5, 5 and 4 base pairs. The central five base pairs (U6·A23 to U10·A19) can be superimposed on the canonical structure with an r.m.s. deviation of only 0.97 Å, whereas at the terminal base pairs (U1·A28 and A14·U15), the backbone atom deviations exceed 3 Å. The r.m.s. deviation between the backbone atoms over the whole molecule is 1.73 Å. The angles between the helical axes of the domains are 13° and 11°. These axes are not coplanar, and the torsion angle between the axes of the two distal helices about the middle helical axis is 70°.

The RNA helix shows a typical ribbon-like structure with a large and hollow minor groove and a deep and narrow major groove. Probably because of one of the kinks, the major groove appears to be pinched between P5 and P18 (2.2 Å) with a compensating extension between P2 and P21 (6.2 Å). The effect on the minor groove is less striking (9.5 to 10.8 Å). Helical parameters reported in Table 1 indicate that the average rise per base pair, base tilt and helical twist are similar to those of canonical RNA. The large propeller twist averaging 18.6°, may be compared with the propeller twists observed in the poly(A) poly(T) stretches of CGCA₆GCG (20.4) (ref. 10) and GCGA₃T₃CGC (20°) (ref. 11) and indicates that the U·A base pairs are readily deformable. The roll angles for the U₅-A₃ steps, averaging 13.8°, are larger than the roll angles for the A₅-U₃ steps (average 7.4°).

Figure 3b shows where the canonical A-RNA structure obtained from fibre diffraction¹ diverges from U(U-A)₆A. Each kink results from torsion angles α (about P-O5') and γ (about C5'-C4') being in the *trans* conformation in only one chain of the duplex instead of the normal *gauche* region. This leads to an extended backbone conformation^{12,13} with phosphorus-phosphorus distances of 6.6 and 6.7 Å, compared with the mean distance of 5.8 Å. The stacking of adjacent bases, which is normally disrupted by this extended conformation, is preserved by the formation of the kink. The kinks are not dependent on the base sequence or the sugar pucker. At position 11 the kink is between U and A, and at position 24 it is between A and U. The ribose at A11 is in the C3'-endo conformation, whereas the sugar at U24 has shifted to the C1'-endo-C2'exo conformation.

The 2'-hydroxyl groups of RNA molecules affect the stereochemistry of the double helix. Model building shows that the C2'-endo conformation is unfavourable for RNA because of steric hindrance, whereas the C3'-endo conformation may be stabilized by hydrogen bonds between the 2'-hydroxyl group and the O4' of the next ribose in the 3' direction. Weak hydrogen bonds of this type (average distance 3.3 Å) occur for about half of the possible interactions in U(U-A)₆A. Similar H-bonds (average distance 3.2 Å) were seen in the helical stems of transfer RNA (refs 2-4). The fibre model shows a distance of 3.4 Å. The likelihood for the formation of these hydrogen bonds is directly related to the rise per residue in the helix¹⁴.

The 2'-hydroxyl groups participate extensively in the formation of hydrogen bonds stabilizing the interhelical interactions between the terminal riboses and exocyclic atoms in the minor groove. These contact regions are close to the kinks, as shown in Fig. 3, and follow the packing behaviour of A-type DNA oligomers¹⁵, where a terminal base-pair stacks against the minor groove of a symmetry-related molecule. The hybrid RNA-DNA structure, r(GCG)d(TATACGC) (ref. 16) also showed a 2'-hydroxyl in an intermolecular contact, but the interaction was through a bridging water molecule.

A different type of contact results from the screw axis about *c*. The major grooves of two related molecules face each other. This interaction is stabilized by van der Waal's contacts between

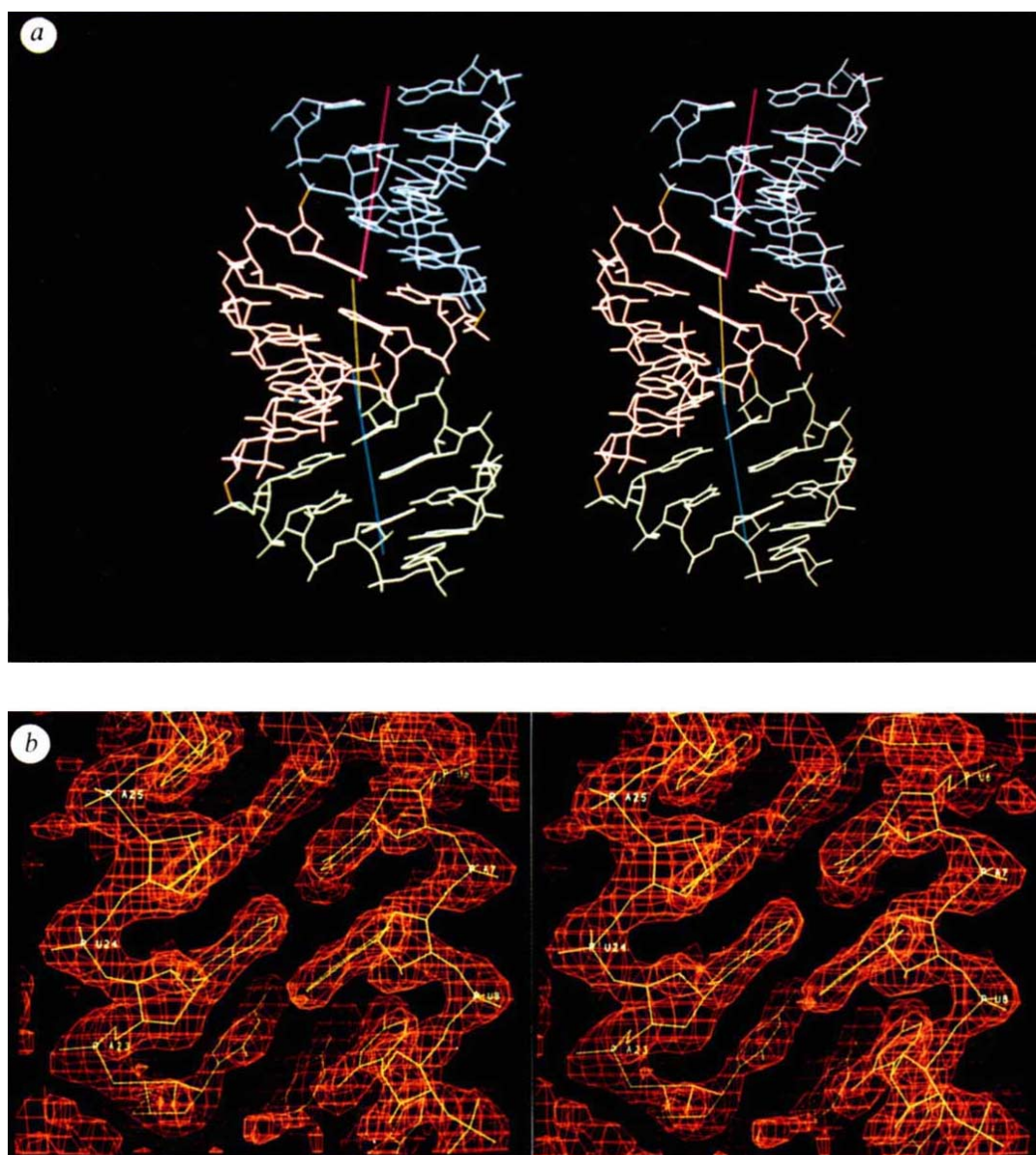


Fig. 2 *a*, Tetradecamer model with the three optimal helical axes corresponding to the three different regions separated by the kinks. The pair U1·A28 is at the top. *b*, Superposition of the last model with the 'omit' electron density map contoured at 20% of its maximum value. A complete diffraction data set to 2.25 Å resolution was collected in the $\Omega/2\theta$ scan mode at room temperature with a CAD4 diffractometer. Out of a total of 3,807 structure factor amplitudes, 2,437 independent reflections were found with $F > 2\sigma(F)$. In the resolution shell between 2.5 and 2.25 Å, 30% of the reflections were still significant ($F > 2\sigma(F)$). The structure was solved by molecular replacement, using the 6-dimensional search program ULTIMA (ref. 29). The starting model was a canonical RNA helix¹ with the weight of the phosphate groups increased by a factor 1.6. The two best solutions were considered (R -factors of 25% and 26% at 10 Å resolution, the third best value was 38%). The model with unweighted phosphates was then introduced, leading to R -values of 42% and 30%. This last position and orientation was the starting point of a least squares procedure. The program CORELS (ref. 30) was first used to position the bases, ribose units and phosphate groups as rigid bodies. In a second step, the Hendrickson-Konnert program, adapted for nucleic acids⁴, was used to carry the refinement to higher resolution. The refinement of the model without solvent molecules converged to an R -factor of 19% for 2,437 reflections, for which $F > 2\sigma(F)$. Solvent molecules were then introduced and the sugar pucker was refined, leading to a final model with 86 water molecules and an R -factor of 13.0%.

U1(C5') and U10(O2P) (3.2 Å) and between U15(C6) and U22(O3') (3.2 Å). The cavity between the major grooves of the neighbouring duplexes is filled with water molecules. This last contact makes the RNA crystal lattice more tightly packed than synthetic A-DNA crystals by eliminating the solvent channels found in the latter.

Crick and Klug predicted that kinks in B-DNA, with angles as large as 90°, would result from the conversion of torsion angle γ to the *trans* conformation¹⁷. Kinks of this type have not been reported. Distortions in DNA helices that have been observed

are actually bends in which the helix curvature is distributed over a series of residues. Two examples are an 'annealed kink' found in a B-DNA oligomer^{18,19} and a bend observed in an A-DNA oligomer GGTATACC (ref. 20). Both of these distortions are different from those observed in $U(UA)_6A$. A true kink has been observed in the DNA fragment complexed to *EcoRI* (ref. 21), but it is clearly stabilized by interactions with the protein. Each kink in $U(UA)_6A$ is stabilized in a different way, as shown in Fig. 3, which emphasizes that the formation of the crystal is a direct consequence of the existence of the kinks.

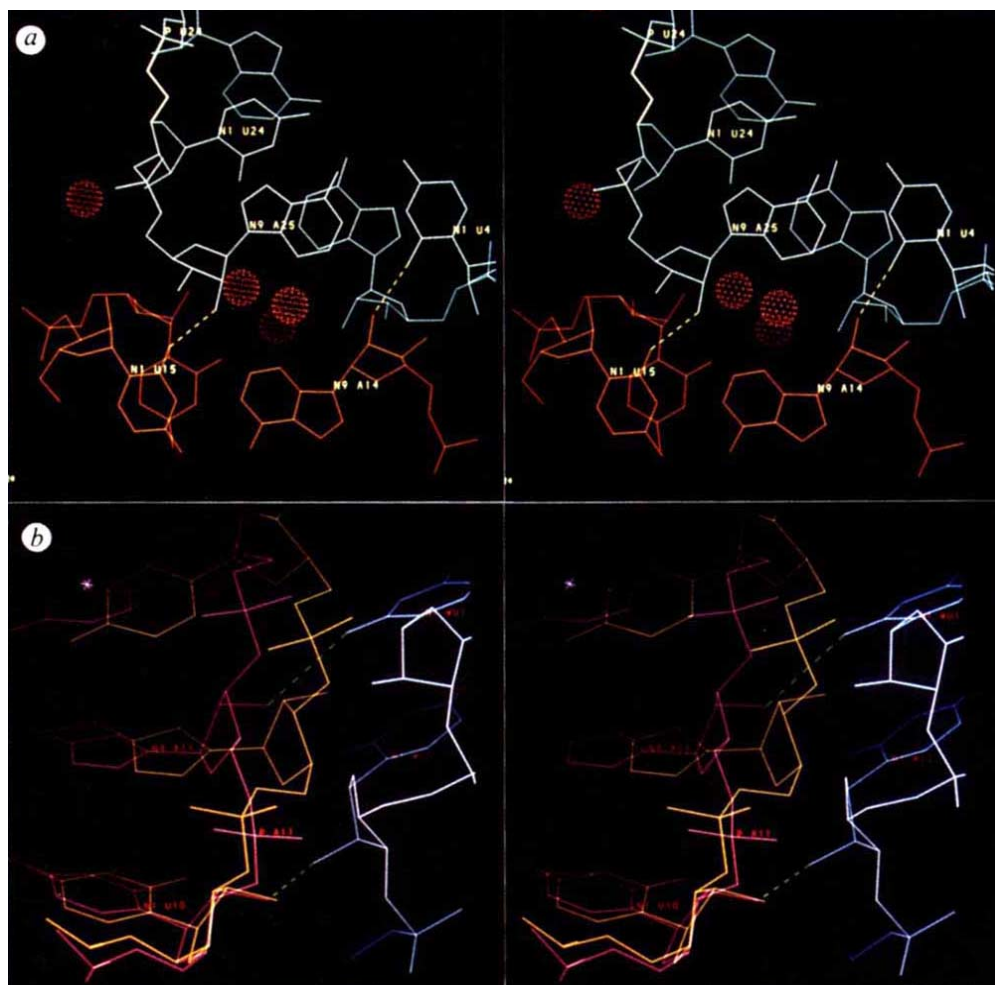


Fig. 3 *a*, Environment of P24 and packing contact between the last base pair A14-U15 and the minor groove of a neighbouring molecule. Note the H-bonds between O2 of U4 and O2' of A14 (distance 2.8 Å), and between O2 of U15 and O2' of A25 (distance 2.6 Å). The site of the conformational change at P24 is in white. At position U24 there are no contacts directly involving the sugars or bases on either side of the unusual P-O5' conformation. This kink appears to be stabilized by an inter-helical hydrogen bond between the 2'-hydroxyl of A25 and O2 from U15 in a symmetry-related helix (2.6 Å). Orange spheres represent solvent molecules. *b*, Kink region near the phosphate group of A11. The figure presents the current structure (pink in the figure) with the fibre model (yellow in the figure). Note that the kink is stabilized by inter-helical H-bonds, between the O2' of U10 and the O2' of a neighbouring U2 (distance 2.4 Å), and between the 2'-hydroxyl of ribose A11 and the O2 atom of U1 in a symmetry-related helix (3.0 Å).

This may explain the high temperature required for crystallizing the RNA in this lattice. At 35 °C the RNA will exist in a variety of conformations that may only be accessible with a low probability at 20 °C. The kink conformation may be trapped in the crystal because of the stabilizing interactions of the packing. The energy required to stabilize such deformations appears to be no more than that corresponding to 2-4 hydrogen bonds. It is clear that although the conformational flexibility of RNA is more restricted than DNA, the RNA helix is not rigid.

The alternating A-U sequence, with its large propeller twist and the alternating high and low roll values, may be more prone to local breaks of the backbone regularity, thus creating a backbone-driven but sequence-dependent recognition feature,

comparable to the polyd(AT) model described by Klug *et al.*²². A direct effect of the kink on the backbone would be to enhance the availability of the 2'-hydroxyl groups. These are the best candidates for specific RNA recognition through their capacity to form directional intermolecular H-bonds.

We thank Dov Rabinovitch for the ULTIMA program. The crystallographic analysis was supported by the CNRS and a grant from the US National Science Foundation. Funds for the synthesis and characterization of the RNA were provided by grants from the National Institutes of Health. J.E.J. acknowledges the CNRS and the Université Louis Pasteur, Strasbourg, France, for a visiting professorship in Strasbourg. We thank Sharon Fateley for her help in the preparation of this manuscript.

Received 27 May; accepted 22 August 1988.

1. Arnott, S., Hukins, D. W. L., Dover, S. D., Fuller, W. & Hodgson, A. R. *J. molec. Biol.* **81**, 107-122 (1973).
2. Sussman, J. L., Holbrook, S. R., Warrant, W. R., Church, G. M. & Kim, S.-H. *J. molec. Biol.* **123**, 607-630 (1978).
3. Hingerty, B., Brown, R. S. & Jack, A. *J. molec. Biol.* **124**, 523-534 (1978).
4. Westhof, E., Dumas, P. & Moras, D. *J. molec. Biol.* **184**, 119-145 (1985).
5. Jaspars, E. M. J. in *Molecular Plant Virology* (ed Davies, J. W.) 155-221 (CRC, Boca Raton, Florida, 1985).
6. Gough, G. R., Collier, K. J., Weith, H. L. & Gilham, P. T. *Nucleic Acids Res.* **7**, 1955-1964 (1979).
7. Gough, G. R., Nadeau, J. G., Gilham, P. T., Singleton, C. K. & Weith, H. L. *Nucleic Acids Res. Symp. No. 7*, 99-102 (1980).
8. Hayes, J. A., Brunden, M. J., Gilham, P. T. & Gough, G. R. *Tetrahedron Lett.* **26**, 2407-2410 (1985).
9. Nadeau, J. G. & Gilham, P. T. *Nucleic Acids Res.* **13**, 8259-8274 (1985).
10. Nelson, H. C. M., Finch, J. T., Luisi, B. F. & Klug, A. *Nature* **330**, 221-226 (1987).
11. Coll, M., Frederick, C. A., Wang, A. H. J. & Rich, A. *Proc. natn. Acad. Sci. U.S.A.* **84**, 8385-8389 (1987).

12. Haran, T. E., Shakked, Z., Wang, A. H. J. & Rich, A. *J. biomol. Struct. Dyn.* **5**, 199-217 (1987).
13. Heineman, U., Lauble, H., Frank, R. & Blöcker, H. *Nucleic Acids Res.* **22**, 9531-9550 (1987).
14. Jack, A., Ladner, J. E. & Klug, A. *J. molec. Biol.* **108**, 619-649 (1976).
15. Wang, A. H. J. & Teng, M. K. *J. Crystal Growth* **90**, 295-310 (1988).
16. Wang, A. H. J. *et al. Nature* **299**, 601-604 (1982).
17. Crick, F. H. C. & Klug, A. *Nature* **255**, 530-533 (1975).
18. Fratini, A. V., Kopka, M. L., Drew, H. R. & Dickerson, R. E. *J. biol. Chem.* **257**, 14686-14707 (1982).
19. Wing, R. *et al. Nature* **287**, 755-758 (1980).
20. Shakked, Z. *et al. J. molec. Biol.* **166**, 183-201 (1983).
21. Frederick, C. A. *et al. Nature* **309**, 327-331 (1984).
22. Klug, A. *et al. J. molec. Biol.* **131**, 669-680 (1979).
23. Shakked, Z. & Rabinovitch, D. *Prog. Biophys. molec. Biol.* **47**, 159-195 (1986).
24. Ohtsuka, E., Tanaka, S. & Ikehara, M. *Nucleic Acids Res.* **1**, 1351-1357 (1974).
25. Ohtsuka, E., Tanaka, S. & Ikehara, M. *Chem. pharmac. Bull.* **25**, 949-959 (1977).
26. Sung, W. L. & Narang, S. A. *Can. J. Chem.* **60**, 111-120 (1982).
27. Gough, G. R., Brunden, M. J. & Gilham, P. T. *Tetrahedron Lett.* **24**, 5321-5324 (1983).
28. Freier, S. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 9373-9377 (1986).
29. Rabinovitch, D. & Shakked, Z. *Acta crystallogr. A* **40**, 195-200 (1984).
30. Sussman, J. L. *Meth. Enzym.* **115**, 271-303 (1985).

Chromatography of carbohydrates

R.R. Townsend, M. Hardy, J.D. Olechno and S.R. Carter

The coupling of high-performance anion-exchange chromatography with pulsed amperometric detection provides a much needed analytical tool for the analysis of the carbohydrates of glycoconjugates.

HIGH-PERFORMANCE anion-exchange chromatography (HPAE) with pulsed amperometric detection (PAD) has recently been applied to the resolution and detection of linkage and branch positional isomers of both neutral and acidic oligosaccharides of biological importance^{1,2}. The coupling of these two techniques enables complete, single-step separation of most types of oligosaccharides found in glycoconjugates (such as glycolipids or glycoproteins) and detection, without derivatization, in the picomole range (10–100 pmols) of the resolved analytes.

Improvements observed in the separation of 1→2, 1→3, and 1→4 positional isomers of oligosaccharides^{1,2} (compared to other HPLC techniques such as reversed-phase³ and amine-bonded HPLC⁴) can be attributed to the accessibility of the most readily ionizable groups of the "neutral" part of the oligosaccharide chains at alkaline pH ($pH \approx 12$ –14) to the pellicular quaternary amine stationary phase. The stationary phase consists of non-porous polystyrene-divinylbenzene beads (10 μm in diameter) on the surface of which are attached 0.1 μm beads of anion exchange material. Because the microbeads have approximately one-millionth the internal volume of a 10- μm bead, rapid exchange of the analytes and solute with the stationary phase occurs, translating into sharper peaks and improved resolution. The resistance of the beads to extremes of pH permits separation of oligosaccharides under both acidic and alkaline conditions.

Neutral oligosaccharides have been separated easily as their oxyanions at alkaline pH¹. Acidic oligosaccharides have been separated according to their formal negative charges near neutral pH or, in cases in which differing selectivity was needed, at alkaline pH². For most of the reducing oligosaccharides obtained from glycolipids and glycoproteins, the most readily ionizable hydroxyl groups are the reducing anomeric hydroxyl, the 2-OH's of unsubstituted pyranoses, and the 3-OH's of 2-acetamido-2-deoxypyranoses^{5,6}. Correlation of retention times with the structure of different oligosaccharides suggests that the accessibility of these readily ionizable hydroxyls to the quaternary amine stationary phase is a major determinant of the enhanced resolution¹.

Sensitive detection of carbohydrates during chromatography has been achieved by derivatization with radiolabels⁷,

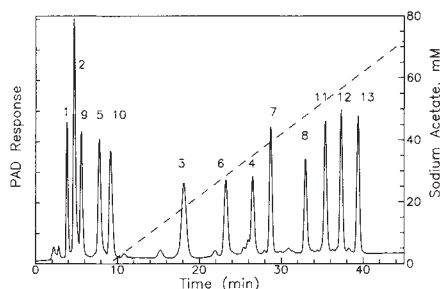


Fig. 1 Chromatography of neutral oligosaccharides using HPAE-PAD. Dashed line indicates sodium acetate gradient.

chromophores⁸, or fluorophores⁹. The development of pulsed electrochemical methods, which bypass the problem of electrode fouling, now enable practical, sensitive detection (10–100 pmol) of underivatized carbohydrates⁹.

In pulsed amperometry, an analytical voltage is maintained for a short time. During this time, a small amount (< 1 per cent) of the analyte is adsorbed to the electrode and oxidized (Johnson, D.C., personal communication). The current generated by the oxidation of the analyte is monitored, and the potential of the

working electrode is then increased to a strongly oxidizing voltage which in effect cleans the surface of the electrode by oxidation of adsorbed analytes to water soluble compounds. Signal is not collected during this or subsequent steps, and the entire pulse sequence is completed in less than one second.

A systematic determination of the molar electrochemical response factors (compared to glucose) have revealed a range from 1.0 to 5.2 for oligosaccharides with up to 18 monosaccharide units¹⁰. In most cases, closely related structures have the same response factor and predictable rules are emerging which should enable the quantification of oligosaccharides in complex mixtures by electrochemical detection.

Chromatography resulting from coupling HPAE with PAD is shown in Fig. 1. Synthetic neutral oligosaccharides which mimic commonly found *N*-linked oligosaccharides of glycoproteins (Table 1) were easily separated. A striking feature of this type of chromatography is that the retention time does not simply increase with the size of the oligosaccharide. For

Table 1 Structures of neutral oligosaccharides separated using HPAE-PAD

Peak	Oligosaccharide structure	Peak	Oligosaccharide structure
1	Fuc α (1,3)GlcNAc β (1,2)Man		Fuc α (1,3)
2	Gal β (1,4)GlcNAc β (1,2)Man Fuc α (1,3)	10	Gal β (1,4)GlcNAc β (1,2)Man α (1,6)Man Gal β (1,4)GlcNAc β (1,2)Man α (1,3)Fuc α (1,3)
3	Gal β (1,4)GlcNAc β (1,3)Gal β (1,4)Glc		Fuc α (1,3)
4	Gal β (1,3)GlcNAc β (1,3)Gal β (1,4)Glc	11	Gal β (1,4)GlcNAc β (1,2)Man α (1,6)Man Gal β (1,4)GlcNAc β (1,2)Man α (1,3)Gal β (1,4)GlcNAc β (1,4)
5	Gal β (1,4)GlcNAc β (1,2)Man-OH Gal β (1,4)GlcNAc β (1,4)		Gal β (1,4)GlcNAc β (1,6)
	Gal β (1,4)GlcNAc β (1,2)Man	12	Gal β (1,4)GlcNAc β (1,2)Man α (1,6)Man Gal β (1,4)GlcNAc β (1,2)Man α (1,3)
6	Gal β (1,4)GlcNAc β (1,4)		Gal β (1,4)GlcNAc β (1,6)
	Gal β (1,4)GlcNAc β (1,6)Man α (1,2)Man	13	Gal β (1,4)GlcNAc β (1,2)Man α (1,6)Man Gal β (1,4)GlcNAc β (1,2)Man α (1,3)Gal β (1,4)GlcNAc β (1,4)
7	Gal β (1,4)GlcNAc β (1,3)		
8	Gal β (1,4)GlcNAc β (1,2)Man α (1,6)Man Gal β (1,4)GlcNAc β (1,2)Man α (1,3)	Oligosaccharides were prepared by the following individuals: Compounds 1, 2, 7, 9, and 10 from Hans Lönn, Department of Organic Chemistry, University of Stockholm; 5, 6, 8, 11–13 from J. Lönngren, Department of Organic Chemistry, University of Stockholm. The oligosaccharides were made available by Y.C. Lee.	
9	Fuc α (1,3)GlcNAc β (1,2)Man α (1,6)Man Fuc α (1,3)GlcNAc β (1,2)Man α (1,3)		

example, peak 10, a nonasaccharide, has approximately the same retention time as a trisaccharide (peak 1). The much shorter retention time of the nonasaccharide compared to its non-fucosylated heptasaccharide (peak 8) was attributed to both fucosyl substitution at the 3-OH of the *N*-acetylglucosamine residues and steric interference of the 2-OH groups of the fucosyl with the acetamido groups of the *N*-acetylglucosamine residues¹. Other features include a significant difference in retention time as a result of the removal of the potential to form an oxyanion at the anomeric hydroxyl (compare reduced peak 5 with reducing peak 6), and complete separation of both linkage positional isomers (peaks 3 and 4) and branch positional isomers (peaks 11 and 12).

Because certain carbohydrates are susceptible to base-catalysed reactions, analytical chromatography under alkaline conditions could yield additional peaks resulting from epimerization of reducing sugars or elimination reactions, particularly those of 3-*O* substituted oligosaccharides. For oligosaccharides with reducing mannose and glucose terminal residues, such as those shown in Table 1, epimerization at room temperature at a pH value of 13 was too slow to produce artefactual peaks. The very small peaks in Fig. 1 were not from base-catalysed reactions. In oligosaccharides with reducing *N*-acetylglucosamine terminal residues, approximately 5 per cent are epimerized to *N*-acetylmannosamine by elution with 0.1 M NaOH. Two methods minimize this complication without loss of column selectivity: chemical reduction of the oligosaccharide, and lowering the NaOH concentration to approximately 30 mM. Glycopeptides and alditols are quite stable under these conditions and can readily be prepared in milligram quantities for nuclear magnetic resonance (NMR) and mass spectrometric analysis. □

R. Reid Townsend and Mark R. Hardy are at the Department of Biology and the McCollum-Pratt Institute, Johns Hopkins University, Baltimore, Maryland 21218, USA. Joseph D. Olechno and Steven R. Carter are at Dionex Corporation, 501 Mercury Drive, Sunnyvale, California 94088, USA. For more information, fill in reader service number 100.

- Hardy, M.R. & Townsend, R.R. *Proc. natn. Acad. Sci. U.S.A.* **85**, 3289-3293 (1988).
- Townsend, R.R., Hardy, M.R., Hinds Gaul, O., & Lee, Y.C. *Analyt. Biochem.*, in the press.
- Tomiya, N., et al. *Analyt. Biochem.* **171**, 73-90 (1988).
- Blanken, W.M., Bergh, M.L.F., Koppen, P.L. & Van den Eijnden, D.H. *Analyt. Biochem.* **145**, 322-330 (1985).
- Rendlemen, J.A. in *Carbohydrates in Solution, Advances in Chemistry Series* Vol. 117 (ed. Gould, R.F.) 51-69 (American Chemical Society, Washington, DC, 1971).
- Neuberger, A. & Wilson, B.M. *Carbohydr. Res.* **17**, 89-95 (1971).
- Takasaki, S., Mizuuchi, T., & Kobata, A. in *Methods in Enzymology* Vol. 83 (ed. Ginsburg) 263-268 (Academic, New York, 1982).
- Wang, W.T., LeDonne, Jr., N.C., Ackerman, B., & Sweeley, C.C. *Analyt. Biochem.* **141**, 366-381 (1984).
- Hughes, S. & Johnson, D.C. *Analytica chim. Acta* **132**, 11-22 (1981).
- Townsend, R.R. & Hardy, M.R., in preparation.

Chemists in L.A.

Thousands of chemists will flock to Los Angeles, California next week for the annual meeting of the American Chemical Society. Over 250 companies will be exhibiting their products; selections from their booths are described below.

SEPRACOR will be introducing its new model MSX-500 laboratory-scale **membrane-mediated solvent extraction system** in booths 1022 and 1024 (*Reader Service No. 101*). Conventional solvent extraction technology relies on the physical mixing of two immiscible phases, which often results in the formation of intractable emulsions and phase entrainment. Sepracor says its system uses a thin, high-surface-area membrane at the liquid/liquid interface to mediate solvent transport, eliminating the need for phase mixing. The benchtop MSX-500 is designed for the development of commercial membrane-based solvent extraction applications, which can later be transferred to larger preparative systems. It consists of a hollow-fibre membrane module flanked by two identical fluid management systems—one for each liquid phase—and may be used with most solvents.

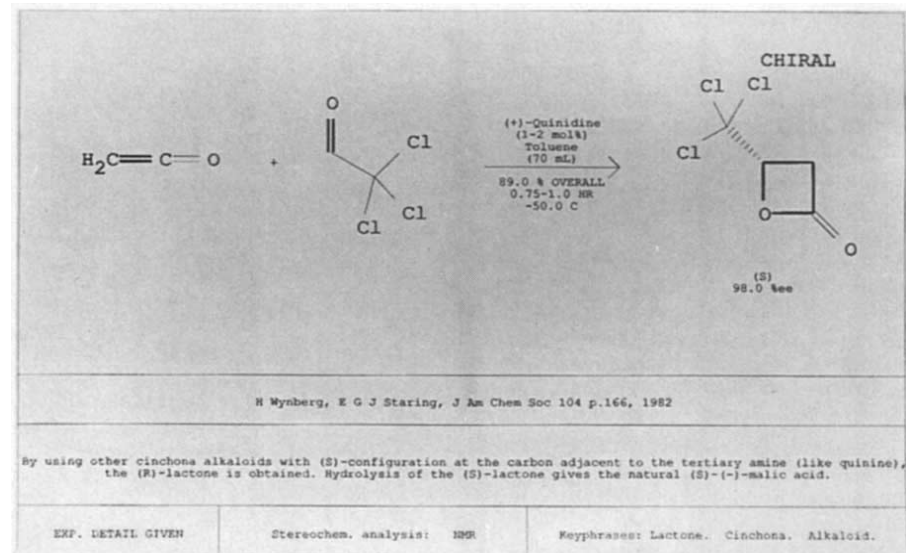
The BioFlow III **benchtop microbial fermenter** has recently been added to New Brunswick Scientific's line of fermentation equipment (*Reader Service No. 102*). The BioFlow III fermenter has an interactive microprocessor which controls the bioreactor temperature, agitation speed, pH, and level of dissolved oxygen, along with the addition of antifoaming agents and nutrients. The instrument will operate in either batch mode or as a chemostat, and is available with vessel sizes of 1.25, 2.5 and 5 litres. The BioFlow III is equipped with RS232 and RS422

communications ports, and can be linked up with dataloggers or a computer running fermentation supervision software for on-line data analysis, graphics plotting and computation of calculated variables. New Brunswick Scientific says the fermenter yields reproducible K_L values and oxygen transfer rates of up to 350 mmol O_2 per litre per hour. The company is exhibiting in booths 522 and 524.

Computer chemistry

Molecular Design, exhibiting in booths 411-415 and 510-514, is developing an **asymmetric synthesis database** named CHIRAS for use with its REaction ACCess System (REACCS) (*Reader Service No. 103*). CHIRAS will be a computer-searchable collection of reactions which result in predominantly one chiral form of a product, and is expected to be of significant value to researchers who need to find the best method for producing organic molecules with controlled absolute and relative stereochemistry. The initial release of CHIRAS, scheduled for the end of this year, will include approximately 6,000 reactions. Another 6,000 reactions will be added by the end of 1989, bringing CHIRAS up to date with all of the reactions included in the chemical literature published between 1975 and 1987. Yearly updates will cover all successive years.

Chemical Design Ltd will have information on its MITIE **supercomputing system**



Example drawn from the CHIRAS database from Molecular Design, showing a reaction which gives a lactone enriched in the (S) enantiomer.

for **molecular modelling** in booths from 405 to 506 (*Reader Service No. 104*). The company says its MITIE system can now be used to calculate the energy of molecular systems and to determine accurate molecular geometries, through quantum mechanics. The computer-intensive technique is only possible because of the intrinsic speed of the T800 transputer chip contained in the MITIE, and its parallel architecture. Chemical Design Ltd says an entry-level MITIE system has a throughput eight times that of a Digital Equipment Corporation VAX 8600 in typical quantum chemistry applications. The MITIE system, including a graphics terminal and Chemical Design's Chem-X modelling software, costs £85,000 (UK).

Hewlett-Packard has recently introduced what it says is the first **benchtop mass spectroscopy workstation** based on the UNIX operating system — the model HP 59940A MS ChemStation (*Reader Service No. 105*). As the data system for the HP 5970B-based benchtop gas chromatography-mass spectroscopy system, the HP 59940A runs the HP-UX operating system, which conforms to AT&T's UNIX System Interface Definition Issue 2, allowing compatibility with a broad range of commercially available software. The multi-tasking capability of the HP 59940A MS ChemStation also allows it to be networked with other systems. When incorporated into the HP-5970B-based GC/MS system, the MS ChemStation offers the capability to customize system menus, and its windowing feature means chemists can simultaneously analyse GC/MS runs, perform library searches, and print reports. Hewlett-Packard is exhibiting in booths 218 and 220.

Sample handling

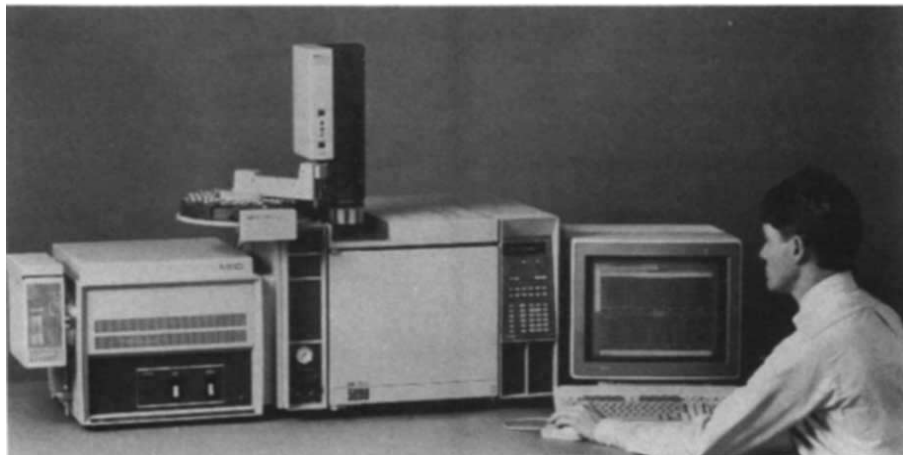
Disposable **Pellet Pestle mixers** for resuspending precipitates in 1.5 ml microtubes are now available from Kontes, exhibiting in booth 1015 (*Reader Service No. 106*). The autoclavable mixers are



Pellet Pestles from Kontes are just the right size for breaking up the precipitate pellets of small samples.

moulded from blue polypropylene and fit the contours of most microtubes, says Kontes. The Pellet Pestles occupy 0.65 ml of tube volume. Because most 1.5 ml microtubes can hold up to 1.7 ml, the precipitates can be resuspended in up to 1 ml of buffer. A cordless motor designed for use with the Pellet Pestle mixers is also available.

Gilson Medical Electronics, in booths 304 and 306, will be showing off its new **ASTED sample preparation system**, based on the principles of dialysis and trace enrichment (*Reader Service No. 107*). ASTED stands for Automated Sequential Trace Enrichment of Dialysates, and the company says the ASTED system separates low molecular weight analytes from complex sample matrices. The ASTED operates on-line with an HPLC system, but interrupt/restart procedures allow the operator to intervene at any time. Gilson says the typical three minute duration of the ASTED sample preparation is generally shorter than the chromatographic analysis, and does not add to the time needed for analysis.



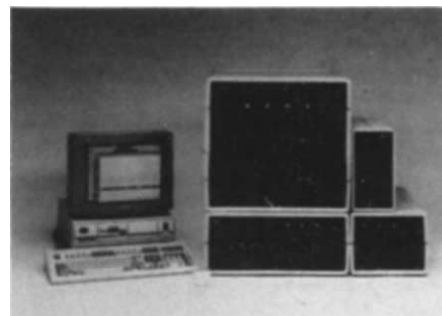
The MS ChemStation (left) from Hewlett-Packard in use with an HP GC/MS (right).

ASTED automatically purges its dialyser and regenerates its trace enrichment cartridge between each sample, and both can be used for 1,000 cycles, says Gilson.

HPLC and columns

Glyco-Pak high pressure liquid chromatography columns for the isolation of **oligosaccharides derived from glycoconjugates** are being introduced by the Waters division of Millipore Corporation, exhibiting in booths 242 and 244 (*Reader Service No. 108*). Glyco-Pak DEAE columns use ion exchange chromatography for the separation of acidic oligosaccharides, and Glyco-Pak N columns operate in a liquid/liquid partition mode to isolate neutral oligosaccharides. The Glyco-Pak columns are polymer-based for greater stability without the contamination encountered with silica-based columns, says Waters.

One of the newest items in the Dionex booths, numbers 736 and 738, is the 4500i series of **liquid chromatographs**, designed



From the Dionex 4500i range of liquid chromatography instruments.

for analysing compounds which are difficult to separate and detect by traditional methods (*Reader Service No. 109*). Dionex says the new systems combine isocratic and quaternary gradient IC and HPLC capability over the entire pH range with unique detection systems and column technology. The chromatographs feature a new family of bipolar-pulsed conductivity and amperometric detection systems and polymer-based pellicular columns. A new type of chromatography module incorporating two valves minimizes dead volume between the 4500i series system components. The 4500i series can be operated unattended with the Microsoft Windows-based Autolon 450 data and control station.

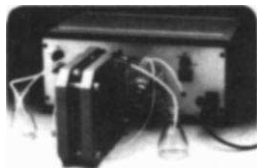
Lee Scientific, exhibiting in booth 521, and Sievers Research have announced the availability of the Sievers model 300 **sulfur chemiluminescence detector** for supercritical fluid chromatography (*Reader Service No. 110*). The model 300 sulfur-selective detector features low-to mid-picogram detection limits for numerous aliphatic mercaptans, sulfides, disulfides, and other organosulfur compounds, including sulfur-containing pesticides and

pharmaceuticals, says Lee Scientific. The company says further that the linear dynamic range for the organosulfur compounds is tested over three orders of magnitude, and selectivity is greater than 10^5 for these compounds over straight-chain hydrocarbons. The detector's response results from the chemiluminescence reaction between analytes containing sulfur-carbon bonds and a reagent gas. The detector output is directly proportional to the quantity of analyte eluting from the SFC column.

ADVERTISEMENTS

LIPOSOMES

REPRODUCIBLE — HOMOGENEOUS —
EXTREMELY RAPID



Our new LIPOSOMAT guarantees rapid (5ml within 60 minutes) preparation of uniformly sized unilamellar liposomes. Liposome size is selectable between 25 and ca. 600 nm diameter. Sample volume is normally 5-10ml or up to 200ml if used in combination with an additional instrument. Lipid concentration can be up to 300 mg/ml.

DIANORM

POB 126, D-8000 Munich 65, FRG

Reader Service No.48

Medwire Div.

stocks a wide variety
of

**TEFLON
COATED***

Platinum alloys,
stainless steel,
silver, multistrand
and single strand
available for prompt
shipment. Write for
descriptive brochure
and price lists.

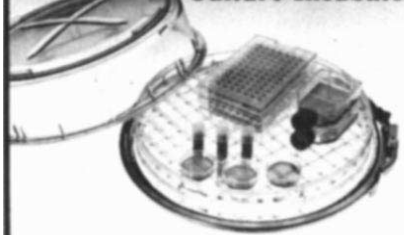
**Medwire Div.**

Affiliate of Sigmund Cohn Corp.

123 So. Columbus Ave.
Mt. Vernon, N.Y. 10553
914-664-5300

*Reg. T.M. of E.I. DuPont

Reader Service No.49

**Economical Modular Tissue
Culture Incubator**

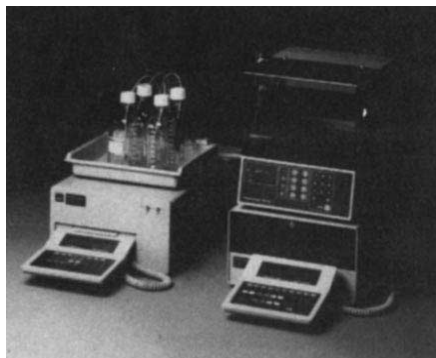
• Safe for AIDS research • Reliable for In-Vitro fertilization
• Versatile enough for thousands of lab uses

Easy to use — Just flush with gas mixture, seal and place at
approp. temp. Perform multiple experiments simultaneously
w/out cross contamination. Thousands in use worldwide.

For information contact: **Flow Laboratories, Ltd. or,**
Billups-Rothenberg, Inc. • P.O. Box 977
Del Mar, CA 92014-0977 USA • (619) 755-3309

Reader Service No.25

Perkin-Elmer has a new **development diode array detector** for HPLC which it says exhibits high sensitivity and excellent



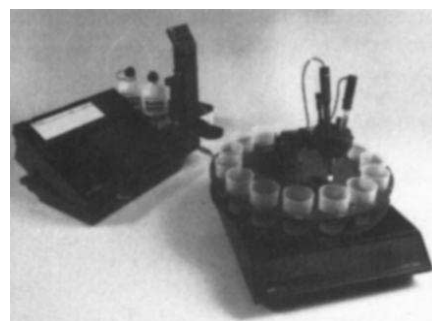
Perkin-Elmer's LC-480 Autoscan is a development diode array detector.

flow cell characteristics with high spectral resolution at low analyte levels — the LC-480 Autoscan (*Reader Service No. 111*). The LC-480 Autoscan can operate as a stand-alone detector offering multichannel detection and arithmetic signal combinations to perform purity checks or spectral suppression techniques. Spectra can be acquired automatically or manually and can be plotted in various formats on a printer-plotter. When coupled with a PC, chromatograms can be transferred in the form of entire UV spectra, and can be manipulated through overlaying, differentiation, subtraction or library functions. Perkin-Elmer is exhibiting in booths 729-733 and 828-832.

An 18-crown-6-type chiral crown ether acts as the chiral selector in the Daicel Crownpak CR **chiral HPLC column** from J.T. Baker (*Reader Service No. 112*). The column is designed for the resolution of amino acids or other compounds which contain a primary amino group near the chiral centre, such as amines and aminophenols. Chiral recognition is achieved when a complex is formed with the sample in an ammonium ion form. Because D-isomers are eluted faster than L-isomers, the absolute configuration of the sample can be ascertained by examining the order of the eluted compounds. The columns are provided with technical literature outlining the conditions required for analysis and a selection of chromatograms of eluted amino acids, amines and aminophenols. J.T. Baker is exhibiting in booths 622 and 624.

Ion analyses

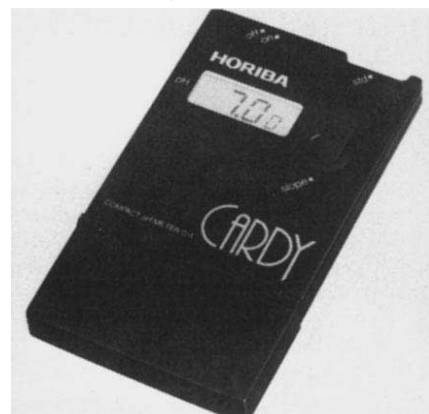
Orion Research, exhibiting in booth 228 and 230, has recently announced the availability of its model 960SC **automatic sample changer**, for repetitive titrations and ion analyses (*Reader Service No. 113*). The sample changer can accommodate up to 20 samples per tray, and trays can be switched by lifting up the one in use and setting a new one in place. There are no



A sample changer for ion analysis from Orion.

hardware connections to the base, and rinse procedures are user-selectable. To prevent carryover from the sensing electrodes between samples, the sample changer can use up to five tray positions for rinsing, with each rinse lasting up to 180 seconds if necessary. The model 960SC uses standard laboratory electrodes with no need for special adaptors or fittings, says the company. Sample weight, identification number, or position number can be entered into the system before the analyses begin.

Horiba Instruments, in booth 133, has a flat-format **pH meter the size of a credit card** with a replaceable electrode (*Reader Service No. 114*). The meter is less than 9



Horiba's credit card-sized pH meter.

mm thick, and the reading precision is 0.01 pH units on an LCD display, says the company. The flat surface of the electrode makes it useful for liquids of as little as 50 µl in volume.

Setting it straight

A review of the CHEF-DR II megabase DNA electrophoresis system from Bio-Rad which appeared in the 18 August issue (*Nature* 334, 636; 1988) incorrectly stated the base range of the system. The system works with DNA sequences of between 2,000 and 10,000,000 bases in length. □

These notes are compiled by Carol Ezzell from information provided by the manufacturers. To obtain further details about these products, use the reader service card found inside the journal. Prices quoted are sometimes nominal, and apply only within the country indicated.